

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 02/090497 A2(51) International Patent Classification⁵: C12N

(21) International Application Number: PCT/US02/14207

(22) International Filing Date: 3 May 2002 (03.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/288,904 4 May 2001 (04.05.2001) US(71) Applicants (for all designated States except US): RE-
NESSEN LLC [US/US]; 3000 Lakeside Drive, Suite 300
- South, Bannockburn, IL 60015 (US). MONSANTO
COMPANY [-US]; 800 N. Lindbergh Boulevard, St.
Louis, MO 63167 (US).

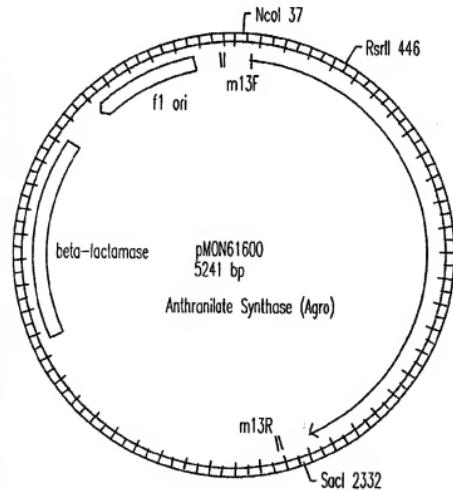
(71) Applicants and

(72) Inventors: WEAVER, Lisa, M. [US/US]; 312 Summer
Haven Court, O'Fallon, MO 63366 (US). LIANG, Jihong
[US/US]; 936 Wellesley Place Drive, Chesterfield, MO
63017 (US). CHEN, Ridong [CN/US]; 11131-I West-
port Station Drive, Maryland Heights, MO 63043 (US).
JEONG, Soon, Seog [KR/US]; 12443-B Whisper Hollow
Drive, Maryland Heights, MO 63043 (US). MITSKY,
Timothy [US/US]; 2262 Rule Avenue, Maryland
Heights, MO 63034 (US). SLATER, Steven [US/US];
21 Brucewood Road, Acton, MA 01720 (US). RAPP,
William [US/US]; 800 North Lindbergh Boulevard, St.
Louis, MO 63167 (US).(74) Agents: STEFFEY, Charles, E. et al.; Schwegman, Lund-
berg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN
55402 (US).*[Continued on next page]*

(54) Title: TRANSGENIC HIGH TRYPTOPHAN PLANTS



WO 02/090497 A2



(57) Abstract: The present invention provides a method for conferring tolerance to an amino acid analog of tryptophan to a plant and/or altering the tryptophan content of a plant by introducing and expressing an isolated DNA segment encoding an anthranilate synthase in the cells of the plant. Transgenic plants transformed with an isolated DNA segment encoding an anthranilate synthase, as well as human or animal food, seeds and progeny derived from these plants, are also provided.



(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, IIR, IIU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PII, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM).

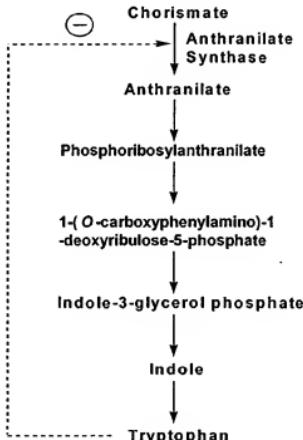
For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TRANSGENIC HIGH TRYPTOPHAN PLANTS

Background of the Invention

The seeds of a number of important crops, including soybean and maize
 5 do not contain sufficient quantities of several amino acids to be nutritionally complete. These amino acids include, but are not limited to: tryptophan, isoleucine, valine, arginine, lysine, methionine and threonine. Therefore, the biosynthetic pathways for these amino acids, and/or biosynthetic pathways for metabolites that feed into those pathways, are potential targets for manipulation
 10 in order to increase the amino acid content of these plants.

Anthranoate synthase (AS, EC 4.1.3.27) catalyzes the first reaction branching from the aromatic amino acid pathway to the biosynthesis of tryptophan in plants, fungi, and bacteria.



The most common form of anthranilate synthase (for example, the maize anthranilate synthase) is a heterotetrameric enzyme consisting of two subunits, the α or TrpE subunit and the β or TrpG subunit. Two α subunits and two β subunits assemble to form the heterotetrameric anthranilate synthases.

20 "Monomeric" forms of AS have also been discovered that comprise a single

polypeptide chain having the activities of both TrpE and TrpG subunits (for example *Rhizobium meliloti*). While monomeric anthranilate synthases comprise just one type of polypeptide, the enzymatically active form of a monomeric anthranilate synthase is typically a homodimer consisting of two such monomeric polypeptides. Both heterotetrameric and monomeric anthranilate synthases catalyze the formation of anthranilate in a reaction utilizing glutamine and chorismate. The domain found on the α subunit (referred to herein as the “ α domain”) binds chorismate and eliminates the enolpyruvate side chain, and the domain found on the β -subunit (referred to herein as the “ β domain”) transfers an amino group from glutamine to the position on the chorismate phenyl ring that resides between the carboxylate and the enolpyruvate moieties.

The next reaction in the synthesis of tryptophan is the transfer of the phosphoribosyl moiety of phosphoribosyl pyrophosphate to anthranilate. The indole ring is formed in two steps involving an isomerization converting the ribose group to a ribulose followed by a cyclization reaction to yield indole glycerol phosphate. The final reaction in the pathway is catalyzed by a single enzyme that may contain either one or two subunits. The reaction accomplishes the cleavage of indole glyceraldehyde-3-phosphate and condensation of the indole group with serine (Umbarger, *Ann. Rev. Biochem.*, **47**, 555 (1978)).

Metabolite flow in the tryptophan pathway in higher plants and microorganisms is apparently regulated through feedback inhibition of anthranilate synthase by tryptophan. Tryptophan may block the conformational rearrangement that is required to activate the β -domain and to create a channel for passage of ammonia toward the active site of the α -domain. Such feedback inhibition by tryptophan is believed to depress the production of tryptophan by anthranilate synthase. See Li J. & Last, R.L. The *Arabidopsis thaliana trp5* mutant has a feedback-resistant anthranilate synthase and elevated soluble tryptophan. *Plant Physiol.* **110**, 51–59 (1996).

Several amino acid residues have been identified as being involved in the feedback regulation of the anthranilate synthase complex from *Salmonella typhimurium*. Such information provides evidence of an amino-terminal regulatory site. *J. Biol. Chem.* **266**, 8328–8335 (1991). Niyogi et al. have

further characterized the anthranilate synthase from certain plants employing a molecular approach. See Niyogi and Fink (Plant Cell, **4**, 721 (1992)) and Niyogi et al. (Plant Cell, **5**, 1011 (1993)). They found that the α -subunits of the 5 *Arabidopsis* anthranilate synthase are encoded by two closely related, nonallelic genes that are differentially regulated. One of these α -subunit genes, ASA1, is induced by wounding and bacterial pathogen infiltration, implicating its involvement in a defense response, whereas the other α -subunit gene, ASA2, is expressed at constitutive basal levels. Both predicted proteins share regions of homology with bacterial and fungal anthranilate synthase proteins, and contain 10 conserved amino acid residues at positions that have been shown to be involved in tryptophan feedback inhibition in bacteria (Caligiuri et al., J. Biol. Chem., **266**, 8328 (1991)).

Amino acid analogs of tryptophan and analogs of the intermediates in the tryptophan biosynthetic pathway (e.g., 5-methyltryptophan, 4-methyltryptophan, 15 5-fluorotryptophan, 5-hydroxytryptophan, 7-azatryptophan, 3 β -indoleacrylic acid, 3-methylanthranilic acid), have been shown to inhibit the growth of both prokaryotic and eukaryotic organisms. Plant cell cultures can be selected for resistance to these amino acid analogs. For example, cultured tobacco, carrot, potato, corn and *Datura innoxia* cell lines have been selected that are resistant to 20 growth inhibition by 5-methyltryptophan (5-MT), an amino acid analog of tryptophan, due to expression of an altered anthranilate synthase.

Ranch et al. (Plant Physiol., **71**, 136 (1983)) selected for 5-MT resistance in cell cultures of *Datura innoxia*, a dicot weed, and reported that the resistant cell cultures contained increased tryptophan levels (8 to 30 times higher than the 25 wild type level) and an anthranilate synthase with less sensitivity to tryptophan feedback inhibition. Regenerated plants were also resistant to 5-MT, contained an altered anthranilate synthase, and had greater concentrations of free tryptophan (4 to 44 times) in the leaves than did the leaves of the control plants. In contrast to the studies with *N. tabacum*, where the altered enzyme was not 30 expressed in plants regenerated from resistant cell lines, these results indicated that the amino acid overproduction phenotype could be selected at the cellular level and expressed in whole plants regenerated from the selected cells in *Datura innoxia*.

Hibberd et al. (U.S. Patent No. 4,581,847, issued April 15, 1986) described 5-MT resistant maize cell lines that contained an anthranilate synthase that was less sensitive to feedback inhibition than wild-type anthranilate synthase. One 5-MT resistant cell line accumulated free tryptophan at levels 5 almost twenty-fold greater than that of non-transformed cell lines.

P. C. Anderson et al. (U.S. Pat. No. 6,118,047) disclose the use of a tryptophan-insensitive α -domain of anthranilate synthase from C28 maize in a transgene to prepare transgenic maize plants (*Zea mays*) exhibiting elevated levels of free tryptophan in the seed(s).

10 Although it is possible to select for 5-MT resistance in certain cell cultures and plants, this characteristic does not necessarily correlate with the overproduction of free tryptophan in whole plants. Additionally, plants regenerated from 5-MT resistant lines frequently do not express an altered form of the enzyme. Nor is it predictable that this characteristic will be stable over a 15 period of time and will be passed along as a heritable trait.

Anthranilate synthase has also been partially purified from crude extracts of cell cultures of higher plants (Hankins et al., *Plant Physiol.*, **57**, 101 (1976); Widholm, *Biochim. Biophys. Acta*, **320**, 217 (1973)). However, it was found to be very unstable. Thus, there is a need to provide plants with a source of 20 anthranilate synthase that can increase the tryptophan content of plants.

Summary of the Invention

The present invention provides nucleic acids encoding an anthranilate synthase (AS) that can be used to generate transgenic plants. When such 25 anthranilate synthase nucleic acids are expressed in a transgenic plant, elevated levels of tryptophan can be achieved within the cells of the plant. In one embodiment, the invention is directed to DNA molecules that encode a monomeric anthranilate synthase, where such a monomeric anthranilate synthase is a natural or genetically engineered chimeric fusion of the α - and β -domains of an anthranilate 30 synthase. The anthranilate synthase gene from a few species (e.g., some bacteria and other microbes) naturally gives rise to a monomeric anthranilate synthase that constitutes a single polypeptide chain. However, most species have a heterotetrameric anthranilate synthase composed of two α and two β domains found

on separate subunits. The invention also contemplates formation of chimeric anthranilate synthase fusion proteins comprising any anthranilate synthase α -domain linked to any β -domain.

In general, the sequence identity of naturally occurring monomeric anthranilate synthases with most plant anthranilate synthases is quite low. 5 However, according to the invention, such monomeric anthranilate synthases can provide high levels of tryptophan when expressed in a plant, despite a low sequence identity with the plant's endogenous anthranilate synthase enzyme. Accordingly, the invention provides monomeric anthranilate synthases that can have divergent 10 sequences and that are capable of efficiently providing high levels of tryptophan in a plant host. For example, transgenic soybean plants containing the monomeric *Agrobacterium tumefaciens* anthranilate synthase can produce from up to about 10,000 to about 12,000 ppm tryptophan in seeds, with average trp levels ranging up to about 7,000 to about 8,000 ppm. In contrast, non-transgenic soybean plants 15 normally have up to only about 100 to about 200 ppm tryptophan in seeds.

Accordingly, the invention provides an isolated DNA sequence encoding a monomeric anthranilate synthase, wherein the monomeric anthranilate synthase has an anthranilate α -domain and an anthranilate β -domain and wherein the monomeric anthranilate synthase is expressed in a plant. Such expression can elevate the level 20 of L-tryptophan in the plant.

The monomeric anthranilate synthase can be naturally monomeric. Examples of organisms from which naturally monomeric anthranilate synthase nucleic acids may be isolated, include but are not limited to organisms such as *Agrobacterium tumefaciens*, *Rhizobium meliloti* (e.g., Genbank Accession No. GI 25 95177), *Mesorhizobium loti* (e.g., Genbank Accession No. GI 13472468), *Brucella melitensis* (e.g., Genbank Accession No. GI 17982357), *Nostoc sp.* PCC7120 (e.g., Genbank Accession Nos. GI 17227910 or GI 17230725), *Azospirillum brasiliense* (e.g., Genbank Accession No. GI 1174156) and *Anabaena* M22983 (e.g., Genbank Accession No. GI 152445). In some embodiments, the isolated DNA encodes an 30 *Agrobacterium tumefaciens* anthranilate synthase having, for example, an amino acid sequence having SEQ ID NO:4 or a nucleotide sequence having any one of SEQ ID NO:1 or 75.

Alternatively, the monomeric anthranilate synthase can be a fusion of any available anthranilate synthase α and β domain. Such α and β domains can be derived from from *Zea mays*, *Ruta graveolens*, *Sulfolobus solfataricus*, *Salmonella typhimurium*, *Serratia marcescens*, *Escherichia coli*, *Agrobacterium tumefaciens*, 5 *Arabidopsis thaliana*, *Rhizobium meliloti* (e.g., Genbank Accession No. GI195177), *Mesorhizobium loti* (e.g., Genbank Accession No. GI 13472468), *Brucella melitensis* (e.g., Genbank Accession No. GI 17982357), *Nostoc sp.* PCC7120 (e.g., Genbank Accession No. GI 17227910 or GI 17230725), *Azospirillum brasilense* (e.g., Genbank Accession No. GI 1174156) and *Anabaena* M22983 (e.g., Genbank 10 Accession No. GI 152445)), soybean, rice, cotton, wheat, tobacco or any gene encoding a subunit or domain of anthranilate synthase. For example, nucleic acids encoding such an α or β domain can be obtained by using the sequence information in any of SEQ ID NO:1 -70, 75-103.

In another embodiment, the invention provides an isolated DNA encoding an 15 α domain of anthranilate synthase from *Zea mays* that comprises SEQ ID NO:5, or SEQ ID NO:66. Such an isolated DNA can have nucleotide sequence SEQ ID NO:2, 67 or 68. The isolated DNA can be operably linked to a promoter and, when expressed in a plant can provide elevated levels of L-tryptophan in the plant.

The isolated DNA can also encode a mutant anthranilate synthase, or a 20 mutant anthranilate synthase domain. Such a mutant anthranilate synthase, or domain thereof, can have one or more mutations. As is known to one of skill in the art, mutations can be silent, can give rise to variant gene products having enzymatic activity similar to wild type or can give rise to derivative gene products that have altered enzymatic acitivity. The invention contemplates all such mutations.

25 The mutated isolated DNA can be generated from a wild type anthranilate synthase nucleic acid either *in vitro* or *in vivo* and can encode, for example, one or more amino acid substitutions, deletions or insertions. Mutant isolated DNAs that generate a mutant anthranilate synthase having increased activity, greater stability, or less sensitivity to feedback inhibition by tryptophan or tryptophan analogs are 30 desirable. In one embodiment, the anthranilate synthase, or a domain thereof, is resistant to inhibition by endogenous L-tryptophan or by tryptophan analogs. For example, the anthranilate synthase can have one or more mutations in the tryptophan-binding pocket or elsewhere that reduces the sensitivity of the

anthranilate synthase, or the domain thereof, to tryptophan inhibition. Among the amino acid residues contemplated for mutation are residues, for example, at about positions 48, 51, 52, 293 and 298. For example, the mutation can be:

- a) at about position 48 replace Val with Phe;
- 5 b) at about position 48 replace Val with Tyr;
- c) at about position 51 replace Ser with Phe;
- d) at about position 51 replace Ser with Cys;
- e) at about position 52 replace Asn with Phe;
- f) at about position 293 replace Pro with Ala;
- 10 g) at about position 293 replace Pro with Gly; or
- h) at about position 298 replace Phe with Trp;

wherein the position of the mutation is determined by alignment of the amino acid sequence of the selected anthranilate synthase with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence. Examples of 15 anthranilate synthases having such mutations include those with SEQ ID NO:58-65, 69, 70, 84-94.

The isolated DNA can encode other elements and functions. Any element or function contemplated by one of skill in the art can be included. For example, the isolated DNA can also include a promoter that can function in a 20 plant cell that is operably linked to the DNA encoding the anthranilate synthase. The isolated DNA can further encode a plastid transit peptide. The isolated DNA can also encode a selectable marker or a reporter gene. Such a selectable marker gene can impart herbicide resistance to cells of said plant, high protein content, high oil content, high lysine content, high isoleucine content, high 25 tocopherol content and the like. The DNA sequence can also comprise a sequence encoding one or more of the insecticidal proteins derived from *Bacillus thuringiensis*.

The invention further provides vectors comprising an isolated DNA of the invention. Such vectors can be used to express anthranilate synthase 30 polypeptides in prokaryotic and eukaryotic cells, to transform plant cells and to generate transgenic plants.

The invention also provides a transgenic plant comprising an isolated DNA of the invention. Expression of these isolated DNAs in the transgenic

plant can result in an elevated level of L-tryptophan, preferably free L-tryptophan, in the transgenic plant, e.g., in the seeds or other parts of the plant. The level is increased above the level of L-tryptophan in the cells of a plant that differ from the cells of the transgenic soybean plant by the absence of the DNA, 5 e.g., the corresponding untransformed cells or an untransformed plant with the same genetic background. The DNA is preferably heritable in that it is preferably transmitted through a complete normal sexual cycle of the fertile plant to its progeny and to further generations.

Transgenic plants that can have such an isolated DNA include 10 dicotyledonous plants (dicots), for example, soybean or canola. Alternatively, the transgenic plants can be monocotyledonous plants (monocots), for example, maize, rice, wheat, barley or sorghum.

The invention also provides a seed of any of the transgenic plants containing any of the isolated DNAs, anthranilate synthase polypeptides, 15 transgenes or vectors of the invention.

The invention further provides an animal feed or human food that contains at least a portion of a plant having an isolated DNA of the invention. Portions of plants that can be included in the animal feed or human food include, for example, seeds, leaves, stems, roots, tubers, or fruits. Desirable portions of 20 plants have increased levels of tryptophan provided by expression of an anthranilate synthase encoded by an isolated DNA of the invention.

The invention further provides a method for altering, preferably increasing, the tryptophan content of a plant (dicot or a monocot) by introducing an isolated DNA of the invention into regenerable cells of the plant. The DNA sequence is preferably operably linked to at least one promoter operable in the 25 plant cells. The transformed cells are identified or selected, and then regenerated to yield a plant comprising cells that can express a functional anthranilate synthase polypeptide. In some embodiments, the DNA encoding the anthranilate synthase, or domain thereof, is a mutant DNA. The introduced 30 DNA is preferably heritable and the plant is preferably a fertile plant. For example, the introduced DNA preferably can be passed by a complete sexual cycle to progeny plants, and can impart the high tryptophan phenotype to subsequent generations of progeny.

- The anthranilate synthase-encoding DNAs, are preferably incorporated into vectors or "transgenes" that can also include DNA sequences encoding transit peptides, such as plastid transit peptides, and selectable marker or reporter genes, operably linked to one or more promoters that are functional in cells of the target plant. The promoter can be, for example, an inducible promoter, a tissue specific promoter, a strong promoter or a weak promoter. Other transcription or translation regulatory elements, e.g., enhancers or terminators, can also be functionally linked to the anthranilate synthase-encoding DNA segment.
- Cells in suspension culture or as embryos, intact tissues or organs can be transformed by a wide variety of transformation techniques, for example, by microprojectile bombardment, electroporation and *Agrobacterium tumefaciens*-mediated transformation, and other procedures available to the art.

Thus, the cells of the transformed plant comprise a native anthranilate synthase gene and a transgene or other DNA segment encoding an exogenous anthranilate synthase. The expression of the exogenous anthranilate synthase in the cells of the plant can lead to increased levels of tryptophan and its secondary metabolites. In some embodiments, such expression confers tolerance to an amount of endogenous L-tryptophan analogue, for example, so that at least about 10% more anthranilate synthase activity is present than in a plant cell having a wild type or tryptophan-sensitive anthranilate synthase.

The invention also provides a method for altering the tryptophan content in a plant comprising: (a) introducing into regenerable cells of a plant a transgene comprising an isolated DNA encoding an anthranilate synthase domain and a plastid transit peptide, operably linked to a promoter functional in the plant cell to yield transformed cells; and (b) regenerating a transformed plant from said transformed plant cells wherein the cells of the plant express the anthranilate synthase domain encoded by the isolated DNA in an amount effective to increase the tryptophan content in said plant relative to the tryptophan content in an untransformed plant of the same genetic background. The domain can be an anthranilate synthase α -domain. The anthranilate synthase domain can have one or more mutations, for example, mutations that reduce the sensitivity of the domain to tryptophan inhibition. Such mutations can be, for example, in the

tryptophan-binding pocket. Such a domain can be, for example, an anthranilate synthase domain from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton or *Zea mays*. *Ruta graveolens* has its own chloroplast transport sequence that may be used with the anthranilate synthase transgene. Accordingly, one of skill in the art may not need to add a plastid transport sequence when using a *Ruta graveolens* DNA.

The present invention also provides novel isolated and purified DNA molecules comprising a DNA encoding a monomeric anthranilate synthase, or a domain thereof. Such an anthranilate synthase DNA can provide high levels of tryptophan when expressed within a plant. In some embodiments, the 15 anthranilate synthase is substantially resistant to inhibition by free L-tryptophan or an analog thereof. Examples of novel DNA sequences contemplated by the invention include but are not limited to DNA molecules isolated from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, or *Zea mays* (maize) or other such anthranilate 20 synthases.

These DNA sequences include synthetic or naturally-occurring 25 monomeric forms of anthranilate synthase that have the α -domain of anthranilate synthase linked to at least one other anthranilate synthase domain on a single polypeptide chain. The monomeric anthranilate synthase can, for example, be a fusion of an anthranilate synthase α or β domain. Such an anthranilate synthase α or β domain can be derived from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton,

wheat, tobacco or *Zea mays* (maize) or any gene encoding a subunit or domain of anthranilate synthase. Such anthranilate synthases and domains thereof are also exemplified herein by the anthranilate synthase nucleic acids isolated from *Agrobacterium tumefaciens*, (SEQ ID NO:1, 75, 84-94), *Zea mays*, (SEQ ID NO:2, 67, 68, 96), *Ruta graveolens* (SEQ ID NO:3), *Anabaena* M22983, *Arabidopsis thaliana* (SEQ ID NO:45), *Azospirillum brasiliense* (SEQ ID NO:78), *Brucella melitensis* (SEQ ID NO:79), *Mesorhizobium loti* (SEQ ID NO:77), *Nostoc sp.* PCC7120 (SEQ ID NO:80 or 81), *Rhizobium meliloti* (SEQ ID NO:7), *Rhodopseudomonas palustris* (SEQ ID NO:57), *Sulfolobus solfataricus* (SEQ ID NO:8), rice (SEQ ID NO:94 or 95), wheat (SEQ ID NO:97), or tobacco (SEQ ID NO:98). These nucleotide sequences encode anthranilate synthases or α -domains thereof from *Agrobacterium tumefaciens* (SEQ ID NO:4, 58-65, 69, 70,); *Zea mays* (SEQ ID NO:5, 66 or 101) and *Ruta graveolens* (SEQ ID NO:6), *Anabaena* M22983, *Azospirillum brasiliense* (SEQ ID NO:78), *Brucella melitensis* (SEQ ID NO:79), *Mesorhizobium loti* (SEQ ID NO:77), *Nostoc sp.* PCC7120 (SEQ ID NO:80 or 81), *Rhizobium meliloti* (SEQ ID NO:7 or 43), *Rhodopseudomonas palustris* (SEQ ID NO:57 or 82), *Sulfolobus solfataricus* (SEQ ID NO:8 or 44), rice (SEQ ID NO:99 or 100), wheat (SEQ ID NO:102), or tobacco (SEQ ID NO:103).

The invention also provides an isolated DNA molecule comprising a DNA sequence encoding an *Agrobacterium tumefaciens* anthranilate synthase or a domain thereof having enzymatic activity. Such a DNA molecule can encode an anthranilate synthase having SEQ ID NO:4, 58-65, 69 or 70, a domain or variant thereof having anthranilate synthase activity. The DNA molecule can also have a sequence comprising SEQ ID NO:1, 75, 84-94, or a domain or variant thereof. Coding regions of any DNA molecule provided herein can also be optimized for expression in a selected organism, for example, a selected plant or microbe. An example of a DNA molecule having optimized codon usage for a selected plant is an *Agrobacterium tumefaciens* anthranilate synthase DNA molecule having SEQ ID NO:75.

The invention also provides an isolated and purified DNA molecule comprising a DNA sequence encoding a *Zea mays* anthranilate synthase domain. Such a DNA molecule can encode an anthranilate synthase domain having SEQ

ID NO:5, 66 or a variant or derivative thereof having anthranilate synthase activity. The DNA molecule can also have a sequence comprising SEQ ID NO:2, 67 or 68, or a domain or variant thereof.

- The invention further provides an isolated DNA molecule of at least 8 nucleotides that hybridizes to the complement of a DNA molecule comprising any one of SEQ ID NO:1, 75 or 84-94 under stringent conditions. Such a DNA molecule can be a probe or a primer, for example, a nucleic acid having any one of SEQ ID NO:9-42 or 47-56. Alternatively, the DNA it can include up to an entire coding region for a selected anthranilate synthase, or a domain thereof.
- Such a DNA can also include a DNA sequence encoding a promoter operable in plant cells and/or a DNA sequence encoding a plastid transit peptide. The invention further contemplates vectors for transformation and expression of these types of DNA molecules in plants and/or microbes.

Functional anthranilate synthase DNA sequences and functional anthranilate synthase polypeptides that exhibit 50%, preferably 60%, more preferably 70%, even more preferably 80%, most preferably 90%, e.g., 95% to 99%, sequence identity to the DNA sequences and amino acid sequences explicitly described herein are also within the scope of the invention. For example, 85% identity means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred.

Alternatively and preferably, two polypeptide sequences are homologous, as this term is used herein, if they have an alignment score of more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M.O., in *Atlas of Protein Sequence and Structure*, 1972, volume 5, National Biomedical Research Foundation, pp. 101-110, and Supplement 2 to this volume, pp. 1-10. The two sequences or parts thereof are more preferably homologous if their amino acids are greater than or equal to 50% identical when optimally aligned using the ALIGN program.

The invention further provides expression vectors for generating a transgenic plant with high seed levels of tryptophan comprising an isolated DNA

sequence encoding a monomeric anthranilate synthase comprising an anthranilate synthase α -domain linked to an anthranilate synthase β -domain and a plastid transit peptide, operably linked to a promoter functional in a plant cell. Such a monomeric anthranilate synthase can, for example, be an *Agrobacterium tumefaciens*,
5 *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc sp.* PCC7120, *Azospirillum brasiliense* or *Anabaena* M22983 anthranilate synthase. The monomeric anthranilate synthase can also be a fusion of anthranilate synthase α and β domains derived from *Agrobacterium tumefaciens*,
10 *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Mesorhizobium loti*, *Nostoc sp.* PCC7120, *Rhizobium meliloti*, *Rhodopseudomonas palustris*, *Ruta graveolens*, *Sulfolobus solfataricus*,
15 *Salmonella typhimurium*, *Serratia marcescens*, soybean, rice, cotton, wheat, tobacco *Zea mays*, or any gene encoding a subunit or domain of anthranilate synthase.

15 The transmission of the isolated and purified anthranilate synthase DNA providing increased levels of tryptophan can be evaluated at a molecular level, e.g., Southern or Northern blot analysis, PCR-based methodologies, the biochemical or immunological detection of anthranilate synthase, or by phenotypic analyses, i.e., whether cells of the transformed progeny can grow in
20 the presence of an amount of an amino acid analog of tryptophan that inhibits the growth of untransformed plant cells.

The invention also provides a method of producing anthranilate synthase in a prokaryotic or eukaryotic host cell, such as a yeast, insect cell, or bacterium, which can be cultured, preferably on a commercial scale. The method includes
25 the steps of introducing a transgene comprising a DNA segment encoding an anthranilate synthase, or a domain thereof, such as a monomeric anthranilate synthase, comprising at least the α and β anthranilate synthase domains, or functional variant thereof, into a host cell and expressing anthranilate synthase in the host cell so as to yield functional anthranilate synthase or domain thereof. A
30 transgene generally includes transcription and translation regulatory elements, e.g., a promoter, functional in host cell, either of eukaryotic or prokaryotic origin. Preferably, the transgene is introduced into a prokaryotic cell, such as *Escherichia coli*, or a eukaryotic cell, such as a yeast or insect cell, that is known

to be useful for production of recombinant proteins. Culturing the transformed cells can lead to enhanced production of tryptophan and its derivatives, which can be recovered from the cells or from the culture media. Accumulation of tryptophan may also lead to the increased production of secondary metabolites in 5 microbes and plants, for example, indole containing metabolites such as simple indoles, indole conjugates, indole alkaloids, indole phytoalexins and indole glucosinalates in plants.

Anthraniate synthases insensitive to tryptophan have the potential to increase a variety of chorismate-derived metabolites, including those derived 10 from phenylalanine due to the stimulation of phenylalanine synthesis by tryptophan via chorismate mutase. *See* Sichl, D. The biosynthesis of tryptophan, tyrosine, and phenylalanine from chorismate in Plant Amino Acids: Biochemistry and Biotechnology, ed. BK Singh, pp 171-204. Other chorismate-derived metabolites that may increase when feedback insensitive anthraniate 15 synthase s are present include phenylpropanoids, flavonoids, and isoflavonoids, as well as those derived from anthraniate, such as indole, indole alkaloids, and indole glucosinolates. Many of these compounds are important plant hormones, plant defense compounds, chemopreventive agents of various health conditions, and/or pharmacologically active compounds. The range of these compounds 20 whose synthesis might be increased by expression of anthraniate synthase depends on the organism in which the anthraniate synthase is expressed. The invention contemplates synthesis of tryptophan and other useful compounds in a variety of prokaryotic and eukaryotic cells or organisms, including plant cells, microbes, fungi, yeast, bacteria, insect cells, and mammalian cells.

25 Hence, the invention provides a method for producing tryptophan comprising: culturing a prokaryotic or eukaryotic host cell comprising an isolated DNA under conditions sufficient to express a monomeric anthraniate synthase encoded by the isolated DNA, wherein the monomeric anthraniate synthase comprises an anthraniate synthase α domain and a anthraniate synthase β domain, and wherein the conditions sufficient to express a monomeric 30 anthraniate synthase comprise nutrients and precursors sufficient for the host cell to synthesize tryptophan utilizing the monomeric anthraniate synthase.

Examples of useful compounds that may be generated upon expression in a variety of host cells and/or organisms include indole acetic acid and other auxins, isoflavanoid compounds important to cardiovascular health found in soy, volatile indole compounds which act as signals to natural enemies of herbivorous insects in maize, anticarcinogens such as indole glucosinolates (indole-3-carbinol) found in the Cruciferae plant family, as well as indole alkaloids such as ergot compounds produced by certain species of fungi. (Barnes et al., Adv Exp Med Biol., **401**, 87 (1996); Frey et al., Proc Natl Acad Sci., **97**, 14801 (2000); Muller et al., Biol Chem., **381**, 679 (2000); Mantegani et al., Farmacol., **54**, 288 (1999); Zeligs, J Med Food., **1**, 67 (1998); Mash et al., Ann NY Acad Sci., **844**, 274 (1998); Melanson et al., Proc Natl Acad Sci., **94**, 13345 (1997); Broadbent et al., Curr Med Chem., **5**, 469 (1998)).

The present invention also provides an isolated and purified DNA molecule of at least seven nucleotide bases that hybridizes under moderate, and preferably, high stringency conditions to the complement of an anthranilate synthase encoding DNA molecule. Such isolated and purified DNA molecules comprise novel DNA segments encoding anthranilate synthase or a domain or mutant thereof. The mutant DNA can encode an anthranilate synthase that is substantially resistant to inhibition by free L-tryptophan or an amino acid analog of tryptophan. Such anthranilate synthase DNA molecules can hybridize, for example, to an *Agrobacterium tumefaciens*, *Rhodopseudomonas palustris* or *Ruta graveolens* anthranilate synthase, or an α -domain thereof, including functional mutants thereof. When these DNA molecules encode a functional anthranilate synthase or an anthranilate synthase domain, they are termed "variants" of the primary DNA molecules encoding anthranilate synthase, anthranilate synthase domains or mutants thereof. Shorter DNA molecules or oligonucleotides can be employed as primers for amplification of target DNA sequences by PCR, or as intermediates in the synthesis of full-length genes.

Also provided is a hybridization probe comprising a novel isolated and purified DNA segment of at least seven nucleotide bases, which is detectably labeled or which can bind to a detectable label, which DNA segment hybridizes under moderate or, preferably, high stringency conditions to the non-coding strand of a DNA molecule comprising a DNA segment encoding an anthranilate

synthase such as a monomeric anthranilate synthase, or a domain thereof, such as the α -domain, including functional mutants thereof, that are substantially resistant to inhibition by an amino acid analog of tryptophan. Moderate and stringent hybridization conditions are well known to the art, see, for example sections 0.47-9.51 of Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition (1989); see also, Sambrook and Russell, Molecular Cloning: A Laboratory Manual, 3rd Edition (January 15, 2001). For example, stringent conditions are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate (SSC); 0.1% sodium lauryl sulfate (SDS) at 50°C, or (2) employ a denaturing agent such as formamide during hybridization, e.g., 50% formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C. Another example is use of 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 μ g/ml), 0.1% sodium dodecylsulfate (SDS), and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

20

Brief Description of the Figures

Figure 1 is a restriction map of pMON61600.

Figure 2 depicts the translated sequence of the *Agrobacterium tumefaciens* anthranilate synthase DNA sequence (upper sequence) (SEQ ID NO:4) and the translated sequence of the anthranilate synthase DNA sequence from *Rhizobium meliloti* (lower sequence) (SEQ ID NO:7).

Figure 3 is a restriction map of pMON34692.

Figure 4 is a restriction map of pMON34697.

Figure 5 is a restriction map of pMON34705.

Figure 6 (A-B) depicts an anthranilate synthase amino acid sequence alignment comparing the *Agrobacterium tumefaciens* α -domain sequence (SEQ ID NO:4) and the *Sulfolobus solfataricus* α -domain sequence (SEQ ID NO:8).

Figure 7 (A-B) depicts the sequences of the 34 primers (SEQ ID NOs 9-42) used to mutate SEQ ID NO:1. The mutated codons are underlined and the changed bases are in lower case.

Figure 8 depicts a restriction map of plasmid pMON13773.

5 Figure 9 depicts a restriction map of plasmid pMON58044.

Figure 10 depicts a restriction map of plasmid pMON53084.

Figure 11 depicts a restriction map of plasmid pMON58045.

Figure 12 depicts a restriction map of plasmid pMON58046.

Figure 13 depicts a restriction map of plasmid pMON38207.

10 Figure 14 depicts a restriction map of plasmid pMON58030.

Figure 15 depicts a restriction map of plasmid pMON58006.

Figure 16 depicts a restriction map of plasmid pMON58041.

Figure 17 depicts a restriction map of plasmid pMON58028.

Figure 18 depicts a restriction map of plasmid pMON58042.

15 Figure 19 depicts a restriction map of plasmid pMON58029.

Figure 20 depicts a restriction map of plasmid pMON58043.

Figure 21 (A-D) depicts a multiple sequence alignment of monomeric “TrpEG” anthranilate synthases having SEQ ID NO:4 and 43 (derived from *Agrobacterium tumefaciens* and *Rhizobium meliloti*, respectively) with the TrpE (α) and TrpG (β) domains of heterotetrameric anthranilate synthases from *Sulfolobus solfataricus* (SEQ ID NO:44) and *Arabidopsis thaliana* (SEQ ID NO:45). Linker regions are underlined.

Figure 22 is a restriction map of plasmid pMON52214.

Figure 23 is a restriction map of plasmid pMON53901.

25 Figure 24 is a restriction map of plasmid pMON39324.

Figure 25 is a restriction map of plasmid pMON39322.

Figure 26 is a restriction map of plasmid pMON39325.

Figure 27 is a graph depicting free tryptophan levels in soybean seeds transformed with pMON39325. There were five observations from each event.

30 NT represents non-transgenic soybean seed.

Figure 28 is a restriction map of plasmid pMON25997.

Figure 29 is a restriction map of plasmid pMON62000.

Figure 30 depicts the sequence of the truncated trpE gene of *Escherichia coli* EMG2 (K-12 wt F+) (SEQ ID NO:46). The first 30bp and the last 150bp of this trpE nucleic acid are connected by an EcoR1 restriction site. The beginning of the trpG gene follows the trpE stop codon.

5 Figure 31 schematically depicts construction of the in-frame deletion in the *E. coli* trpE gene.

Figure 32 (A-C) depicts the DNA (SEQ ID NO:1) and amino acid (SEQ ID NO:4) sequences of the α -domain of the anthranilate synthase gene isolated from *Agrobacterium tumefaciens*.

10 Figure 33 (A-C) depicts the DNA (SEQ ID NO:2) sequence of the α -domain of the anthranilate synthase gene isolated from *Zea mays*. Figure 33 (D) depicts the amino acid (SEQ ID NO:5) sequence of the α -domain of the anthranilate synthase gene isolated from *Zea mays*.

Figure 34 is a restriction map of plasmid pMON58120.

15 Figure 35 (A-E) provides a sequence comparison of anthranilate synthase amino acid sequences from *Agrobacterium tumefaciens* (AgrTu_15889565) (SEQ ID NO:4), *Rhizobium meliloti* (RhiMe_136328) (SEQ ID NO:7), *Mesorhizobium loti* (MesLo_13472468) (SEQ ID NO:77), *Azospirillum brasilense* (AzoBr_1717765) (SEQ ID NO:78), *Brucella melitensis* (BruMe_17986732) (SEQ ID NO:79), *Nostoc sp.* (Nostoc_17227910) (SEQ ID NO:80), *Nostoc sp.* (Nostoc_17230725) (SEQ ID NO:81), and *Rhodopseudomonas palustris* (RhoPa_TrpEG) (SEQ ID NO:82).

20 Figure 36 (A-B) provides an optimized nucleotide sequence for *Agrobacterium tumefaciens* anthranilate synthase (SEQ ID NO:75).

25 Figure 37 (A-C) provides an alignment of the wild type (top strand) and optimized (bottom strand) *Agrobacterium tumefaciens* anthranilate synthase nucleotide sequences (SEQ ID NO:1 and 75). These two sequences are 94% identical.

30 **Detailed Description of the Invention**

The present invention provides isolated DNAs, vectors, host cells and transgenic plants comprising an isolated nucleic acid encoding an anthranilate synthase capable of providing high levels of tryptophan upon expression within

the plant. In one embodiment, the isolated nucleic acid encodes a monomeric anthranilate synthase (AS). In other embodiments, the isolated nucleic acid encodes an anthranilate synthase, or a domain thereof, that is substantially resistant to inhibition by free L-tryptophan or an amino acid analog of tryptophan. Expression of the anthranilate synthase, or domain thereof, elevates the level of tryptophan, e.g., free tryptophan in the seed, over the level present in the plant absent such expression.

Methods are also provided for producing transgenic plants having nucleic acids associated with increased anthranilate synthase activity, and producing cultured cells, plant tissues, plants, plant parts and seeds that produce high levels of tryptophan. Such transgenic plants can preferably sexually transmit the ability to produce high levels of tryptophan to their progeny. Also described are methods for producing isolated DNAs encoding mutant anthranilate synthases, and cell culture selection techniques to select for novel genotypes that overproduce tryptophan and/or are resistant to tryptophan analogs. For example, to produce soybean lines capable of producing high levels of tryptophan, transgenic soybean cells that contain at least one of the isolated DNAs of the invention, are prepared and characterized, then regenerated into plants. Some of the isolated DNAs are resistant to growth inhibition by the tryptophan analog. The methods provided in the present invention may also be used to produce increased levels of free tryptophan in dicot plants, such as other legumes, as well as in monocots, such as the cereal grains.

Definitions

As used herein, "altered" levels of tryptophan in a transformed plant, plant tissue, plant part or plant cell are levels which are greater or lesser than the levels found in the corresponding untransformed plant, plant tissue, plant part or plant cell.

As used herein, a " α -domain" is a portion of an enzyme or enzymatic complex that binds chorismate and eliminates the enolpyruvate side chain. Such an α -domain can be encoded by a TrpE gene. In some instances, the α -domain is a single polypeptide that functions only to bind chorismate and to eliminate the enolpyruvate side chain from chorismate. In other instances, the α -domain is

part of a larger polypeptide that can carry out other enzymatic functions in addition to binding chorismate and eliminating the enolpyruvate side chain from chorismate.

The term “ β -domain” refers to a portion of an enzyme or enzymatic complex that transfers an amino group from glutamine to the position on the chorismate ring that resides between the carboxylate and the enolpyruvate moieties. Such a β -domain can be encoded by a TrpG gene. In some instances, the β -domain is a single polypeptide that functions only to transfer an amino group from glutamine to the position on the chorismate ring that resides between the carboxylate and the enolpyruvate moieties. In other instances, the β -domain is part of a larger polypeptide that can carry out other enzymatic functions in addition to transferring an amino group from glutamine to the position on the chorismate ring that resides between the carboxylate and the enolpyruvate moieties.

As used herein, “an amino acid analog of tryptophan” is an amino acid that is structurally related to tryptophan and that can bind to the tryptophan-binding site in a wild type anthranilate synthase. These analogs include, but are not limited to, 6-methylanthranilate, 5-methyltryptophan, 4-methyltryptophan, 5-fluorotryptophan, 5-hydroxytryptophan, 7-azatryptophan, 3 β -indoleacrylic acid, 3-methylanthranilic acid, and the like.

The term “consists essentially of” as used with respect to the present DNA molecules, sequences or segments is defined to mean that a major portion of the DNA molecule, sequence or segment encodes an anthranilate synthase. Unless otherwise indicated, the DNA molecule, sequence or segment generally does not encode proteins other than an anthranilate synthase.

The term “complementary to” is used herein to mean that the sequence of a nucleic acid strand could hybridize to all, or a portion, of a reference polynucleotide sequence. For illustration, the nucleotide sequence “TATACT” has 100% identity to a reference sequence 5’-TATAC-3’ but is 100% complementary to a reference sequence 5’-GTATA-3’.

As used herein, an “exogenous” anthranilate synthase is an anthranilate synthase that is encoded by an isolated DNA that has been introduced into a host cell, and that is preferably not identical to any DNA sequence present in the cell

in its native, untransformed state. An "endogenous" or "native" anthranilate synthase is an anthranilate synthase that is naturally present in a host cell or organism.

As used herein, "increased" or "elevated" levels of free L-tryptophan in a plant cell, plant tissue, plant part or plant are levels that are about 2 to 200 times, preferably about 5 to 150 times, and more preferably about 10-100 times, the levels found in an untransformed plant cell, plant tissue, plant part or plant, i.e., one where the genome has not been altered by the presence of an exogenous anthranilate synthase nucleic acid or domain thereof. For example, the levels of free L-tryptophan in a transformed plant seed are compared with those in an untransformed plant seed ("the starting material").

DNA molecules encoding an anthranilate synthase, and DNA molecules encoding a transit peptide or marker/reporter gene are "isolated" in that they were taken from their natural source and are no longer within the cell where they normally exist. Such isolated DNA molecules may have been at least partially prepared or manipulated *in vitro*, e.g., isolated from a cell in which they are normally found, purified, and amplified. Such isolated DNA molecules can also be "recombinant" in that they have been combined with exogenous DNA molecules or segments. For example, a recombinant DNA can be an isolated DNA that is operably linked to an exogenous promoter, or to a promoter that is endogenous to the host cell.

As used herein with respect to anthranilate synthase, the term "monomeric" means that two or more anthranilate synthase domains are incorporated in a functional manner into a single polypeptide chain. The monomeric anthranilate synthase may be assembled *in vivo* into a dimeric form. Monomeric anthranilate synthase nucleic acids and polypeptides can be isolated from various organisms such as *Agrobacterium tumefaciens*, *Anabaena* M22983, *Azospirillum brasilense*, *Brucella melitensis*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120 or *Rhizobium meliloti*. Alternatively, monomeric anthranilate synthase nucleic acids and polypeptides can be constructed from a combination of domains selected from any convenient monomeric or multimeric anthranilate synthase gene. Such organisms include, for example, *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*.

Azospirillum brasiliense, *Brucella melitensis*, *Mesorhizobium loti*, *Nostoc sp.* PCC7120, *Rhizobium meliloti*, *Rhodopseudomonas palustris*, *Ruta graveolens*, *Sulfolobus solfataricus*, *Salmonella typhimurium*, *Serratia marcescens*, soybean, rice, cotton *Zea mays*, or any gene encoding a subunit or domain of anthranilate synthase. Nucleic acids encoding the selected domains can be linked recombinantly. For example, a nucleic acid encoding the C-terminus of an α -domain can be linked to a nucleic acid encoding the N-terminus of the β -domain, or vice versa, by forming a phosphodiester bond. As an alternative, such single domain polypeptides can be linked chemically. For example, the α -domain can be linked via its C-terminus to the N-terminus of the β -domain, or vice versa, by forming a peptide bond.

As used herein, a "native" gene means a gene that has not been changed *in vitro*, i.e., a "wild-type" gene that has not been mutated *in vitro*.

The term "plastid" refers to the class of plant cell organelles that includes amyloplasts, chloroplasts, chromoplasts, elaioplasts, eoplasts, etioplasts, leucoplasts, and proplastids. These organelles are self-replicating, and contain what is commonly referred to as a "chloroplast genome," a circular DNA molecule that ranges in size from about 120 to about 217 kb, depending upon the plant species, and which usually contains an inverted repeat region.

As used herein, "polypeptide" means a continuous chain of amino acids that are all linked together by peptide bonds, except for the N-terminal and C-terminal amino acids that have amino and carboxylate groups, respectively, and that are not linked in peptide bonds. Polypeptides can have any length and can be post-translationally modified, for example, by glycosylation or phosphorylation.

As used herein, a plant cell, plant tissue or plant that is "resistant or tolerant to inhibition by an amino acid analog of tryptophan" is a plant cell, plant tissue, or plant that retains at least about 10% more anthranilate synthase activity in the presence of an analog of L-tryptophan, than a corresponding wild type anthranilate synthase. In general, a plant cell, plant tissue, or plant that is "resistant or tolerant to inhibition by an amino acid analog of tryptophan" can grow in an amount of an amino acid analog of tryptophan that normally inhibits growth of the untransformed plant cell, plant tissue, or plant, as determined by

methodologies known to the art. For example, a homozygous backcross converted inbred plant transformed with a DNA molecule that encodes an anthranilate synthase that is substantially resistant or tolerant to inhibition by an amino acid analog of tryptophan grows in an amount of an amino acid analog of 5 tryptophan that inhibits the growth of the corresponding, i.e., substantially isogenic, recurrent inbred plant.

As used herein, an anthranilate synthase that is "resistant or tolerant to inhibition by tryptophan or an amino acid analog of tryptophan" is an anthranilate synthase that retains greater than about 10% more activity than a 10 corresponding "wild-type" or native susceptible anthranilate synthase, when the tolerant/resistant and wild type anthranilate synthases are exposed to equivalent amounts of tryptophan or an amino acid analog of tryptophan. Preferably the resistant or tolerant anthranilate synthase retains greater than about 20% more activity than a corresponding "wild-type" or native susceptible anthranilate 15 synthase.

As used herein with respect to anthranilate synthase, the term "a domain thereof," includes a structural or functional segment of a full-length anthranilate synthase. A structural domain includes an identifiable structure within the anthranilate synthase. An example of a structural domain includes an alpha 20 helix, a beta sheet, an active site, a substrate or inhibitor binding site and the like.

A functional domain includes a segment of an anthranilate synthase that performs an identifiable function such as a tryptophan binding pocket, an active site or a substrate or inhibitor binding site. Functional domains of anthranilate synthase include those portions of anthranilate synthase that can catalyze one 25 step in the biosynthetic pathway of tryptophan. For example, an α -domain is a domain that can be encoded by *trpE* and that can transfer NH₃ to chorismate and form anthranilate. A β -domain can be encoded by *trpG* and can remove an amino group from glutamine to form ammonia. Hence, a functional domain includes enzymatically active fragments and domains of an anthranilate synthase.

30 Mutant domains of anthranilate synthase are also contemplated. Wild type anthranilate synthase nucleic acids utilized to make mutant domains include, for example, any nucleic acid encoding a domain of *Agrobacterium tumefaciens*, *Anabaena M22983*, *Arabidopsis thaliana*, *Azospirillum brasilense*, *Brucella*

melitensis, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Rhodopseudomonas palustris*, *Ruta graveolens*, *Sulfolobus solfataricus*, *Salmonella typhimurium*, *Serratia marcescens*, soybean, rice, cotton, wheat, tobacco *Zea mays*, or any gene encoding a subunit or domain of anthranilate synthase that can comprise at least one amino acid substitution in the coding region thereof. Domains that are mutated or joined to form a monomeric anthranilate synthase having increased tryptophan biosynthetic activity, greater stability, reduced sensitivity to tryptophan or an analog thereof, and the like, are of particular interest.

10

General Concepts

The present invention relates to novel nucleic acids and methods for obtaining plants that produce elevated levels of free L-tryptophan. The overproduction results from the introduction and expression of a nucleic acid 15 encoding anthranilate synthase, or a domain thereof. Such anthranilate synthase nucleic acids include wild type or mutant α -domains, or monomeric forms of anthranilate synthase. A monomeric form of anthranilate synthase comprises at least two anthranilate synthase domains in a single polypeptide chain, e.g., an α -domain linked to a β -domain.

20

Native plant anthranilate synthases are generally quite sensitive to feedback inhibition by L-tryptophan and analogs thereof. Such inhibition constitutes a key mechanism for regulating the tryptophan synthetic pathway. Therefore, an anthranilate synthase or a domain thereof that is highly active, more efficient or that is inhibited to a lesser extent by tryptophan or an analog 25 thereof will likely produce elevated levels of tryptophan. According to the invention, the *Agrobacterium tumefaciens* anthranilate synthase is particularly useful for producing high levels of tryptophan.

30

To generate high levels of tryptophan in a plant or a selected host cell, the selected anthranilate synthase nucleic acid is isolated and may be manipulated *in vitro* to include regulatory signals required for gene expression in plant cells or other cell types. Because the tryptophan biosynthetic pathway in plants is reported to be present within plastids, the exogenous anthranilate synthase nucleic acids are either introduced into plastids or are modified by adding a

nucleic acid segment encoding an amino-terminal plastid transit peptide. Such a plastid transit peptide can direct the anthranilate synthase gene product into plastids. In some instances the anthranilate synthase may already contain a plastid transport sequence, in which case there is no need to add one.

5 In order to alter the biosynthesis of tryptophan, the nucleic acid encoding an anthranilate synthase activity must be introduced into plant cells or other host cells and these transformed cells identified, either directly or indirectly. An entire anthranilate synthase or a useful portion or domain thereof can be used. The anthranilate synthase is stably incorporated into the plant cell genome. The
10 transcriptional signals controlling expression of the anthranilate synthase must be recognized by and be functional within the plant cells or other host cells. That is, the anthranilate synthase must be transcribed into messenger RNA, and the mRNA must be stable in the plant cell nucleus and be transported intact to the cytoplasm for translation. The anthranilate synthase mRNA must have
15 appropriate translational signals to be recognized and properly translated by plant cell ribosomes. The polypeptide gene product must substantially escape proteolytic attack in the cytoplasm, be transported into the correct cellular compartment (e.g. a plastid) and be able to assume a three-dimensional conformation that will confer enzymatic activity. The anthranilate synthase must
20 further be able to function in the biosynthesis of tryptophan and its derivatives; that is, it must be localized near the native plant enzymes catalyzing the flanking steps in biosynthesis (presumably in a plastid) in order to obtain the required substrates and to pass on the appropriate product.

Even if all these conditions are met, successful overproduction of
25 tryptophan is not a predictable event. The expression of some transgenes may be negatively affected by nearby chromosomal elements. If the high level of tryptophan is achieved by mutation to reduce feedback inhibition, there may be other control mechanisms compensating for the reduced regulation at the anthranilate synthase step. There may be mechanisms that increase the rate of
30 breakdown of the accumulated amino acids. Tryptophan and related amino acids must be also overproduced at levels that are not toxic to the plant. Finally, the introduced trait must be stable and heritable in order to permit commercial development and use.

Isolation and Identification of DNA Coding for an Anthranilate Synthase

Nucleic acids encoding an anthranilate synthase can be identified and isolated by standard methods, for example, as described by Sambrook et al.,
5 Molecular Cloning: A Laboratory Manual, 2nd Edition (1989); Sambrook and Russell, Molecular Cloning: A Laboratory Manual, 3rd Edition (January 15, 2001). For example, a DNA sequence encoding an anthranilate synthase or a domain thereof can be identified by screening of a DNA or cDNA library generated from nucleic acid derived from a particular cell type, cell line, primary
10 cells, or tissue. Examples of libraries useful for identifying and isolating an anthranilate synthase include, but are not limited to, a cDNA library derived from *Agrobacterium tumefaciens* strain A348, maize inbred line B73 (Stratagene, La Jolla, California, Cat. #937005, Clontech, Palo Alto, California, Cat. # FL1032a, #FL1032b, and FL1032n), genomic library from maize inbred
15 line Mo17 (Stratagene, Cat. #946102), genomic library from maize inbred line B73 (Clontech, Cat. # FL1032d), genomic DNA from *Anabaena* M22983 (e.g., Genbank Accession No. GI 152445), *Arabidopsis thaliana*, *Azospirillum brasilense* (e.g., Genbank Accession No. GI 1174156), *Brucella melitensis* (GI 17982357), *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti* (e.g.,
20 Genbank Accession No. GI 13472468), *Nostoc* sp. PCC7120 (e.g., Genbank Accession No. GI 17227910 or GI 17230725), *Rhizobium meliloti* (e.g., Genbank Accession No. GI 95177), *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco *Zea mays* (maize) or other species. Moreover,
25 anthranilate synthase nucleic acids can be isolated by nucleic acid amplification procedures using genomic DNA, mRNA or cDNA isolated from any of these species.

Screening for DNA fragments that encode all or a portion of the sequence encoding an anthranilate synthase can be accomplished by screening plaques
30 from a genomic or cDNA library for hybridization to a probe of an anthranilate synthase gene from other organisms or by screening plaques from a cDNA expression library for binding to antibodies that specifically recognize anthranilate synthase. DNA fragments that hybridize to anthranilate synthase

probes from other organisms and/or plaques carrying DNA fragments that are immunoreactive with antibodies to anthranilate synthase can be subcloned into a vector and sequenced and/or used as probes to identify other cDNA or genomic sequences encoding all or a portion of the desired anthranilate synthase gene.

- 5 Preferred cDNA probes for screening a maize or plant library can be obtained from plasmid clones pDPG600 or pDPG602.

A cDNA library can be prepared, for example, by random oligo priming or oligo dT priming. Plaques containing DNA fragments can be screened with probes or antibodies specific for anthranilate synthase. DNA fragments encoding 10 a portion of an anthranilate synthase gene can be subcloned and sequenced and used as probes to identify a genomic anthranilate synthase gene. DNA fragments encoding a portion of a bacterial or plant anthranilate synthase can be verified by determining sequence homology with other known anthranilate synthase genes or by hybridization to anthranilate synthase-specific messenger RNA. Once cDNA 15 fragments encoding portions of the 5', middle and 3' ends of an anthranilate synthase are obtained, they can be used as probes to identify and clone a complete genomic copy of the anthranilate synthase gene from a genomic library.

Portions of the genomic copy or copies of an anthranilate synthase gene can be sequenced and the 5' end of the gene identified by standard methods 20 including either by DNA sequence homology to other anthranilate synthase genes or by RNAase protection analysis, for example, as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition (1989); Sambrook and Russell, Molecular Cloning: A Laboratory Manual, 3rd Edition (January 15, 2001). The 3' and 5' ends of the target gene can also be located by computer 25 searches of genomic sequence databases using known AS coding regions. Once portions of the 5' end of the gene are identified, complete copies of the anthranilate synthase gene can be obtained by standard methods, including cloning or polymerase chain reaction (PCR) synthesis using oligonucleotide primers complementary to the DNA sequence at the 5' end of the gene. The 30 presence of an isolated full-length copy of the anthranilate synthase gene can be verified by hybridization, partial sequence analysis, or by expression of a maize anthranilate synthase enzyme.

Exemplary isolated DNAs of the invention include DNAs having the following nucleotide SEQ ID NO:

- SEQ ID NO:1 -- *Agrobacterium tumefaciens* (wild type)
SEQ ID NO:2 -- *Zea mays* (wild type)
5 SEQ ID NO:3 -- *Ruta graveolens*
SEQ ID NO:46 -- truncated TrpE gene of *E. coli* EMG2 (K-12 wt
F+)
SEQ ID NO:67 -- *Zea mays* (C28 mutant)
SEQ ID NO:68 -- *Zea mays* (C28 + terminator)
10 SEQ ID NO:71 -- Chloroplast Targeting Peptide (g)
SEQ ID NO:73 -- Chloroplast Targeting Peptide (a)
SEQ ID NO:75 -- *Agrobacterium tumefaciens* (optimized)
SEQ ID NO:76 -- *Rhodopseudomonas palustris*
SEQ ID NO:83 -- *Rhodopseudomonas palustris* (RhoPa_TrpEG)
15 SEQ ID NO:84 -- *Agrobacterium tumefaciens* V48F mutant .
SEQ ID NO:85 -- *Agrobacterium tumefaciens* V48Y mutant
SEQ ID NO:86 -- *Agrobacterium tumefaciens* S51F mutant
SEQ ID NO:87 -- *Agrobacterium tumefaciens* S51C mutant
SEQ ID NO:88 -- *Agrobacterium tumefaciens* N52F mutant
20 SEQ ID NO:89 -- *Agrobacterium tumefaciens* P293A mutant
SEQ ID NO:90 -- *Agrobacterium tumefaciens* P293G mutant
SEQ ID NO:91 -- *Agrobacterium tumefaciens* F298W mutant
SEQ ID NO:92 -- *Agrobacterium tumefaciens* S50K mutant
SEQ ID NO:93 -- *Agrobacterium tumefaciens* F298A mutant
25 SEQ ID NO:94 -- rice
SEQ ID NO:95 -- rice isozyme
SEQ ID NO:96 -- maize (U.S. Patent 6,118,047 to Anderson)
SEQ ID NO:97 -- wheat
SEQ ID NO:98 -- tobacco
30 Certain primers are also useful for the practise of the invention, for example, primers having SEQ ID NO:9-42, 47-56.

The invention also contemplates any isolated nucleic acid encoding an anthranilate synthase having, for example, any one of the following amino acid sequences.

- SEQ ID NO:4 *Agrobacterium tumefaciens* (wild type)
5 SEQ ID NO:5 *Zea mays* (wild type)
SEQ ID NO:6 *Ruta graveolens*
SEQ ID NO:7 *Rhizobium meliloti*
SEQ ID NO:8 *Sulfolobus solfataricus*
SEQ ID NO:43 *Rhizobium meliloti*
10 SEQ ID NO:44 *Sulfolobus solfataricus*
SEQ ID NO:45 *Arabidopsis thaliana*
SEQ ID NO:57 *Rhodopseudomonas palustris*
SEQ ID NO:58 *Agrobacterium tumefaciens* V48F mutant
SEQ ID NO:59 *Agrobacterium tumefaciens* V48Y mutant
15 SEQ ID NO:60 *Agrobacterium tumefaciens* S51F mutant
SEQ ID NO:61 *Agrobacterium tumefaciens* S51C mutant
SEQ ID NO:62 *Agrobacterium tumefaciens* N52F mutant
SEQ ID NO:63 *Agrobacterium tumefaciens* P293A mutant
SEQ ID NO:64 *Agrobacterium tumefaciens* P293G mutant
20 SEQ ID NO:65 *Agrobacterium tumefaciens* F298W mutant
SEQ ID NO:66 *Zea mays* C28 mutant
SEQ ID NO:67 *Agrobacterium tumefaciens* S50K mutant
SEQ ID NO:70 *Agrobacterium tumefaciens* F298A mutant
SEQ ID NO:74 Chloroplast Targeting Peptide (a)
25 SEQ ID NO:72 Chloroplast Targeting Peptide (g)
SEQ ID NO:77 *Mesorhizobium loti* (MesLo_13472468)
SEQ ID NO:78 *Azospirillum brasiliense* (AzoBr_1717765)
SEQ ID NO:79 *Brucella meliensis* (BruMe_17986732)
SEQ ID NO:80 *Nostoc sp.* (Nostoc_17227910)
30 SEQ ID NO:81 *Nostoc sp.* (Nostoc_17230725)
SEQ ID NO:82 *Rhodopseudomonas palustris* RhoPa_TrpEG
SEQ ID NO:99 -- rice
SEQ ID NO:100 -- rice isozyme

SEQ ID NO:101 -- maize (U.S. Patent 6,118,047 to Anderson)

SEQ ID NO:102 -- wheat

SEQ ID NO:103 -- tobacco

Any of these nucleic acids and polypeptides can be utilized in the practice of the
5 invention, as well as any mutant, variant or derivative thereof.

Monomeric Anthranilate Synthases

According to the invention, monomeric anthranilate synthases from plant and non-plant species are functional in plants and can provide high levels of
10 tryptophan. Surprisingly, monomeric anthranilate synthases from non-plant species function very well in plants even though the sequences of these monomeric anthranilate synthases have low homology with most plant anthranilate synthases. For example, monomeric anthranilate synthases from species as diverse as bacteria, protists, and microbes can be used successfully. In particular, monomeric
15 anthranilate synthases from bacterial species such as *Agrobacterium tumefaciens*, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc sp. PCC7120*, *Azospirillum brasilense* and *Anabaena M22983* are functional in plants and can provide high levels of tryptophan, despite the rather low sequence identity of these monomeric anthranilate synthases with most plant anthranilate synthases.

20 Transgenic plants containing, for example, the wild type monomeric *Agrobacterium tumefaciens* anthranilate synthase can produce up to about 10,000 to about 12,000 ppm tryptophan in seeds, with average trp levels ranging up to about 7,000 to about 8,000 ppm. Non-transgenic soybean plants normally have up to only about 100 to about 200 ppm tryptophan in seeds. By comparison transgenic plants
25 containing an added mutant *Zea mays* α domain produce somewhat lower levels of tryptophan (e.g., averages up to about 3000 to about 4000 ppm).

Monomeric enzymes may have certain advantages over multimeric enzymes. For example, while the invention is not to be limited to a specific mechanism, a monomeric enzyme may provide greater stability, coordinated expression, and the
30 like. When domains or subunits of a heterotetrameric anthranilate synthase are synthesized in vivo, those domains/subunits must properly assemble into a heterotetrameric form before the enzyme becomes active. Addition of a single domain of anthranilate synthase by transgenic means to a plant may not provide

overproduction of the entire heterotetrameric enzyme because there may not be sufficient endogenous amounts of the non-transgenic domains to substantially increase levels of the functional tetramer. Hence, nucleic acids, vectors and enzymes encoding a monomeric anthranilate synthase can advantageously be used
5 to overproduce all of the enzymatic functions of anthranilate synthase.

According to the invention, anthranilate synthase domains from species that naturally produce heterotetrameric anthranilate synthases can be fused or linked to provide monomeric anthranilate synthases that can generate high tryptophan levels when expressed within a plant cell, plant tissue or seed. For example, a monomeric
10 anthranilate synthase can be made by fusing or linking the α and β domains of anthranilate synthase so that the sequence of the α - β fusion generally aligns with an anthranilate synthase that is naturally monomeric. Examples of sequence alignments of monomeric and heterotetrameric anthranilate synthases are shown in Figures 21 and 35. Using such sequence alignments, the spacing and orientation of
15 anthranilate synthase domains can be adjusted or modified to generate a monomeric anthranilate construct from heterotetrameric domains that optimally aligns with naturally monomeric anthranilate synthases. Such a fusion protein can be used to increase tryptophan levels in the tissues of a plant.

Heterotetrameric anthranilate synthases, such as the *Sulfolobus solfataricus*
20 anthranilate synthase (e.g., Genbank Accession No. GI1004323), share between about 30% to about 87% sequence homology with heterotetrameric anthranilate synthases from other plant and microbial species. Monomeric anthranilate synthases, such as the *A. tumefaciens* anthranilate synthase, have between about 83% and about 52% identity to the other monomeric enzymes such as *Rhizobium meliloti* (Genbank Accession No. GI 15966140) and *Azospirillum brasiliense* (Genbank Accession No. 1717765), respectively. Bae et al., *Rhizobium meliloti* anthranilate synthase gene: cloning, sequence, and expression in *Escherichia coli*. *J. Bacteriol.* 171, 3471–3478 (1989); De Troch et al., Isolation and characterization of the *Azospirillum brasiliense* trpE(G) gene, encoding anthranilate synthase. *Curr. 25 Microbiol.* 34, 27–32 (1997).

However, the overall sequence identity shared between naturally monomeric and naturally heterotetrameric anthranilate synthases can be less than 30%. Hence, visual alignment rather than computer-generated alignment, may be needed to

optimally align monomeric and heterotetrameric anthranilate synthases. Landmark structures and sequences within the anthranilate synthases can facilitate sequence alignments. For example, the motif "LLES" is part of a β -sheet of the β -sandwich that forms the tryptophan-binding pocket of anthranilate synthases. Such landmark 5 sequences can be used to more confidently align divergent anthranilate synthase sequences, and are especially useful for determination of key residues involved in tryptophan binding.

To accomplish the fusion or linkage of anthranilate synthase domains, the C-terminus of the selected TrpE or α -domain is linked to the N-terminus of the TrpG 10 domain or β -domain. In some cases, a linker peptide may be utilized between the domains to provide the appropriate spacing and/or flexibility. Appropriate linker sequences can be identified by sequence alignment of monomeric and heterotetrameric anthranilate synthases.

The selected β -domains can be cloned, for example, by hybridization, 15 PCR amplification or as described in Anderson et al., U.S. Pat. No. 6,118,047. A plastid transit peptide sequence can also be linked to the anthranilate synthase coding region using standard methods. For example, an *Arabidopsis* small subunit (SSU) chloroplast targeting peptide (CTP, SEQ ID NO:71-74) may be used for this purpose. *See also*, Stark et al., (1992) Science 258: 287. The fused 20 gene can then be inserted into a suitable vector for plant transformation as described herein.

Anthranilate Synthase Mutants

Mutant anthranilate synthases contemplated by the invention can have 25 any type of mutation including, for example, amino acid substitutions, deletions, insertions and/or rearrangements. Such mutants can be derivatives or variants of anthranilate synthase nucleic acids and polypeptides specifically identified herein. Alternatively, mutant anthranilate synthases can be obtained from any available species, including those not explicitly identified herein. The mutants, 30 derivatives and variants can have identity with at least about 30% of the amino acid positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity. In a preferred embodiment, polypeptide derivatives and variants have identity with at least about 50% of the amino acid

positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity. In a more preferred embodiment, polypeptide derivatives and variants have identity with at least about 60% of the amino acid positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity. In a more preferred embodiment, polypeptide derivatives and variants have identity with at least about 70% of the amino acid positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity. In an even more preferred embodiment, polypeptide derivatives and variants have identity with at least about 80% of the amino acid positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity. In an even more preferred embodiment, polypeptide derivatives and variants have identity with at least about 90% of the amino acid positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity. In an even more preferred embodiment, polypeptide derivatives and variants have identity with at least about 95% of the amino acid positions of any one of SEQ ID NO:4-8, 43-45, 57-66, 69-70, 77-82, 99-103 and have anthranilate synthase activity.

In one embodiment, anthranilate synthase mutants, variants and derivatives can be identified by hybridization of any one of SEQ ID NO:1-3, 9-42, 46, 47-56, 67-68, 75-76, 83-98, or a fragment or primer thereof under moderate or, preferably, high stringency conditions to a selected source of nucleic acids. Moderate and stringent hybridization conditions are well known to the art, see, for example sections 0.47-9.51 of Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition (1989); *see also*, Sambrook and Russell, Molecular Cloning: A Laboratory Manual, 3rd Edition (January 15, 2001). For example, stringent conditions are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate (SSC); 0.1% sodium lauryl sulfate (SDS) at 50°C, or (2) employ a denaturing agent such as formamide during hybridization, e.g., 50% formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C. Another example is use of 50%

formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% sodium dodecylsulfate (SDS), and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

The invention further provides hybridization probes and primers comprising a novel isolated and purified DNA segment of at least seven nucleotide bases, which can be detectably labeled or bind to a detectable label. Such a hybridization probe or primer can hybridize under moderate or high stringency conditions to either strand of a DNA molecule that encodes an anthranilate synthase. Examples of such hybridization probes and primers include any one of SEQ ID NO:9-42, 47-56.

The anthranilate synthase can be any anthranilate synthase, or a mutant or domain thereof, such as the α -domain. The anthranilate synthase can be a monomeric anthranilate synthase. Functional mutants are preferred, particularly those that can generate high levels of tryptophan in a plant, for example, those mutants that are substantially resistant to inhibition by an amino acid analog of tryptophan.

Nucleic acids encoding mutant anthranilate synthases can also be generated from any convenient species, for example, from nucleic acids encoding any domain of *Agrobacterium tumefaciens*, *Anabaena* M22983 (e.g., Genbank Accession No. GI 152445), *Arabidopsis thaliana*, *Azospirillum brasilense* (e.g., Genbank Accession No. GI 1174156), *Brucella melitensis* (e.g., Genbank Accession No. GI 17982357), *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti* (e.g., Genbank Accession No. GI 13472468), *Nostoc* sp. PCC7120 (e.g., Genbank Accession No. GI 17227910 or GI 17230725), *Rhizobium meliloti* (e.g., Genbank Accession No. GI 95177), *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco *Zea mays* (maize) or any gene encoding a subunit or domain of anthranilate synthase.

Mutants having increased anthranilate synthase activity, reduced sensitivity to feedback inhibition by tryptophan or analogs thereof, and/or the ability to generate increased amounts of tryptophan in a plant are desirable. Such

mutants do have a functional change in the level or type of activity they exhibit and are sometimes referred to as "derivatives" of the anthranilate synthase nucleic acids and polypeptides provided herein.

However, the invention also contemplates anthranilate synthase variants 5 as well as anthranilate synthase nucleic acids with "silent" mutations. As used herein, a silent mutation is a mutation that changes the nucleotide sequence of the anthranilate synthase but that does not change the amino acid sequence of the encoded anthranilate synthase. A variant anthranilate synthase is encoded by a mutant nucleic acid and the variant has one or more amino acid changes that do 10 not substantially change its activity when compared to the corresponding wild type anthranilate synthase. The invention is directed to all such derivatives, variants and anthranilate synthases nucleic acids with silent mutations.

DNA encoding a mutated anthranilate synthase that is resistant and/or tolerant to L-tryptophan or amino acid analogs of tryptophan can be obtained by 15 several methods. The methods include, but are not limited to:

1. spontaneous variation and direct mutant selection in cultures;
2. direct or indirect mutagenesis procedures on tissue cultures of any cell types or tissue, seeds or plants;
3. mutation of the cloned anthranilate synthase gene by methods such as 20 by chemical mutagenesis; site specific or site directed mutagenesis (Sambrook et al., cited *supra*), transposon mediated mutagenesis (Berg et al., *Biotechnology*, **1**, 417 (1983)), and deletion mutagenesis (Mitra et al., *Molec. Gen. Genetic.*, **215**, 294 (1989));
4. rational design of mutations in key residues; and
- 25 5. DNA shuffling to incorporate mutations of interest into various anthranilate synthase nucleic acids.

For example, protein structural information from available anthranilate synthase proteins can be used to rationally design anthranilate synthase mutants that have a high probability of having increased activity or reduced sensitivity to 30 tryptophan or tryptophan analogs. Such protein structural information is available, for example, on the *Solfidobus solfataricus* anthranilate synthase (Knochel et. al., *Proc. Natl. Acad. Sci. USA*, **96**, 9479-9484 (1999)). Rational design of mutations can be accomplished by alignment of the selected

anthranilate synthase amino acid sequence with the anthranilate synthase amino acid sequence from an anthranilate synthase of known structure, for example, *Sulfolobus solfataricus*. See Figures 6, 21 and 35. The predicted tryptophan binding and catalysis regions of the anthranilate synthase protein can be assigned

5 by combining the knowledge of the structural information with the sequence homology. For example, residues in the tryptophan binding pocket can be identified as potential candidates for mutation to alter the resistance of the enzyme to feedback inhibition by tryptophan. Using such structural information, several *Agrobacterium tumefaciens* anthranilate synthase mutants were rationally

10 designed in the site or domain involved in tryptophan binding.

Using such sequence and structural analysis, regions analogous to the monomeric *Agrobacterium tumefaciens* anthranilate synthase at approximately positions 25-60 or 200-225 or 290-300 or 370-375 were identified in the monomeric *Agrobacterium tumefaciens* anthranilate synthase as being

15 potentially useful residues for mutation to produce active anthranilate synthases that may have less sensitivity to tryptophan feedback inhibition. More specifically, amino acids analogous to P29, E30, S31, I32, S42, V43, V48, S50, S51, N52, N204, P205, M209, F210, G221, N292, P293, F298 and A373 in the monomeric *Agrobacterium tumefaciens* anthranilate synthase are being

20 potentially useful residues for mutation to produce active anthranilate synthases that may have less sensitivity to tryptophan feedback inhibition. The invention contemplates any amino acid substitution or insertion at any of these positions. Alternatively, the amino acid at any of these positions can be deleted.

Site directed mutagenesis can be used to generate amino acid

25 substitutions, deletions and insertions at a variety of sites. Examples of specific mutations made within the *Agrobacterium tumefaciens* anthranilate synthase coding region include the following:

at about position 48 replace Val with Phe (see e.g., SEQ ID NO:58);
at about position 48 replace Val with Tyr (see e.g., SEQ ID NO:59);

30 at about position 51 replace Ser with Phe (see e.g., SEQ ID NO:60);
at about position 51 replace Ser with Cys (see e.g., SEQ ID NO:61);
at about position 52 replace Asn with Phe (see e.g., SEQ ID NO:62);
at about position 293 replace Pro with Ala (see e.g., SEQ ID NO:63);

at about position 293 replace Pro with Gly (see e.g., SEQ ID NO:64); or
at about position 298 replace Phe with Trp (see e.g., SEQ ID NO:65).

- Similar mutations can be made in analogous positions of any anthranilate synthase by alignment of the amino acid sequence of the anthranilate synthase to be mutated with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence. One example of an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence that can be used for alignment is SEQ ID NO:4.

- Useful mutants can also be identified by classical mutagenesis and genetic selection. A functional change can be detected in the activity of the enzyme encoded by the gene by exposing the enzyme to free L-tryptophan or amino acid analogs of tryptophan, or by detecting a change in the DNA molecule using restriction enzyme mapping or DNA sequence analysis.

- For example, a gene encoding an anthranilate synthase substantially tolerant to 5-methyltryptophan can be isolated from a 5-methyltryptophan tolerant cell line. See U.S. Patent No. 4,581,847, issued April 15, 1986, the disclosure of which is incorporated by reference herein. Briefly, partially differentiated plant cell cultures are grown and subcultured with continuous exposures to low levels of 5-methyltryptophan. 5-methyltryptophan concentrations are then gradually increased over several subculture intervals. Cells or tissues growing in the presence of normally toxic 5-methyltryptophan levels are repeatedly subcultured in the presence of 5-methyltryptophan and characterized. Stability of the 5-methyltryptophan tolerance trait of the cultured cells may be evaluated by growing the selected cell lines in the absence of 5-methyltryptophan for various periods of time and then analyzing growth after exposing the tissue to 5-methyltryptophan. Cell lines that are tolerant by virtue of having an altered anthranilate synthase enzyme can be selected by identifying cell lines having enzyme activity in the presence of normally toxic, i.e., growth inhibitor, levels of 5-methyltryptophan.

- The anthranilate synthase gene cloned from a 5-MT- or 6-MA-resistant cell line can be assessed for tolerance to 5-MT, 6-MA, or other amino acid analogs of tryptophan by standard methods, as described in U.S. Patent No. 4,581,847, issued April 15, 1986, the disclosure of which is incorporated by reference herein.

Cell lines with an anthranilate synthase of reduced sensitivity to 5-methyltryptophan inhibition can be used to isolate a 5-methyltryptophan-resistant anthranilate synthase. A DNA library from a cell line tolerant to 5-methyltryptophan can be generated and DNA fragments encoding all or a portion 5 of an anthranilate synthase gene can be identified by hybridization to a cDNA probe encoding a portion of an anthranilate synthase gene. A complete copy of the altered gene can be obtained either by cloning and ligation or by PCR synthesis using appropriate primers. The isolation of the altered gene coding for anthranilate synthase can be confirmed in transformed plant cells by determining 10 whether the anthranilate synthase being expressed retains enzyme activity when exposed to normally toxic levels of 5-methyltryptophan. See, Anderson et al., U.S. Pat. No. 6,118,047.

Coding regions of any DNA molecule provided herein can also be optimized for expression in a selected organism, for example, a selected plant or 15 other host cell type. An example of a DNA molecule having optimized codon usage for a selected plant is an *Agrobacterium tumefaciens* anthranilate synthase DNA molecule having SEQ ID NO:75. This optimized *Agrobacterium tumefaciens* anthranilate synthase DNA (SEQ ID NO:75) has 94% identity with SEQ ID NO:1.

20

Transgenes and Vectors

Once a nucleic acid encoding anthranilate synthase or a domain thereof is obtained and amplified, it is operably combined with a promoter and, optionally, with other elements to form a transgene.

25

Most genes have regions of DNA sequence that are known as promoters and which regulate gene expression. Promoter regions are typically found in the flanking DNA sequence upstream from the coding sequence in both prokaryotic and eukaryotic cells. A promoter sequence provides for regulation of transcription of the downstream gene sequence and typically includes from about 30 50 to about 2,000 nucleotide base pairs. Promoter sequences also contain regulatory sequences such as enhancer sequences that can influence the level of gene expression. Some isolated promoter sequences can provide for gene expression of heterologous genes, that is, a gene different from the native or

homologous gene. Promoter sequences are also known to be strong or weak or inducible. A strong promoter provides for a high level of gene expression, whereas a weak promoter provides for a very low level of gene expression. An inducible promoter is a promoter that provides for turning on and off of gene expression in response to an exogenously added agent or to an environmental or developmental stimulus. Promoters can also provide for tissue specific or developmental regulation. An isolated promoter sequence that is a strong promoter for heterologous genes is advantageous because it provides for a sufficient level of gene expression to allow for easy detection and selection of transformed cells and provides for a high level of gene expression when desired.

- The promoter in a transgene of the invention can provide for expression of anthranilate synthase from a DNA sequence encoding anthranilate synthase. Preferably, the coding sequence is expressed so as to result in an increase in tryptophan levels within plant tissues, for example, within the seeds of the plant.
- 15 In another embodiment, the coding sequence is expressed so as to result in increased tolerance of the plant cells to feedback inhibition or to growth inhibition by an amino acid analog of tryptophan or so as to result in an increase in the total tryptophan content of the cells. The promoter can also be inducible so that gene expression can be turned on or off by an exogenously added agent.
- 20 For example, a bacterial promoter such as the P_{lac} promoter can be induced to varying levels of gene expression depending on the level of isothiopropylgalactoside added to the transformed bacterial cells. It may also be preferable to combine the gene with a promoter that provides tissue specific expression or developmentally regulated gene expression in plants. Many promoters useful in the practice of the invention are available to those of skill in the art.

Preferred promoters will generally include, but are not limited to, promoters that function in bacteria, bacteriophage, plastids or plant cells. Useful promoters include the CaMV 35S promoter (Odell et al., Nature, **313**, 810 (1985)), the CaMV 19S (Lawton et al., Plant Mol. Biol., **9**, 31F (1987)), *nos* (Ebert et al., PNAS USA, **84**, 5745 (1987)), *Adh* (Walker et al., PNAS USA, **84**, 6624 (1987)), sucrose synthase (Yang et al., PNAS USA, **87**, 4144 (1990)), α -tubulin, napin, actin (Wang et al., Mol. Cell. Biol., **12**, 3399 (1992)), *cab*

(Sullivan et al., Mol. Gen. Genet., **215**, 431 (1989)), PEPCase promoter (Hudspeth et al., Plant Mol. Biol., **12**, 579 (1989)), the 7S-alpha'-conglycinin promoter (Beachy et al., EMBO J., **4**, 3047 (1985)) or those associated with the R gene complex (Chandler et al., The Plant Cell, **1**, 1175 (1989)). Other useful 5 promoters include the bacteriophage SP6, T3, and T7 promoters.

Plastid promoters can be also be used. Most plastid genes contain a promoter for the multi-subunit plastid-encoded RNA polymerase (PEP) as well as the single-subunit nuclear-encoded RNA polymerase. A consensus sequence for the nuclear-encoded polymerase (NEP) promoters and listing of specific 10 promoter sequences for several native plastid genes can be found in Hajdukiewicz et al., 1997, EMBO J. Vol. 16 pp. 4041-4048, which is hereby in its entirety incorporated by reference.

Examples of plastid promoters that can be used include the *Zea mays* plastid RRN (ZMRRN) promoter. The ZMRRN promoter can drive expression 15 of a gene when the *Arabidopsis thaliana* plastid RNA polymerase is present. Similar promoters that can be used in the present invention are the Glycine max plastid RRN (SOYRRN) and the Nicotiana tabacum plastid RRN (NTRRN) promoters. All three promoters can be recognized by the *Arabidopsis* plastid RNA polymerase. The general features of RRN promoters are described by 20 Hajdukiewicz et al. and U.S. Patent 6,218,145.

Moreover, transcription enhancers or duplications of enhancers can be used to increase expression from a particular promoter. Examples of such enhancers include, but are not limited to, elements from the CaMV 35S promoter and octopine synthase genes (Last et al., U.S. Patent No. 5,290,924, issued 25 March 1, 1994). For example, it is contemplated that vectors for use in accordance with the present invention may be constructed to include the *ocs* enhancer element. This element was first identified as a 16 bp palindromic enhancer from the octopine synthase (*ocs*) gene of *Agrobacterium* (Ellis et al., EMBO J., **6**, 3203 (1987)), and is present in at least 10 other promoters (Bouchez 30 et al., EMBO J., **8**, 4197 (1989)). It is proposed that the use of an enhancer element, such as the *ocs* element and particularly multiple copies of the element, will act to increase the level of transcription from adjacent promoters when applied in the context of monocot transformation. Tissue-specific promoters,

including but not limited to, root-cell promoters (Conkling et al., Plant Physiol., 93, 1203 (1990)), and tissue-specific enhancers (Fromm et al., The Plant Cell, 1, 977 (1989)) are also contemplated to be particularly useful, as are inducible promoters such as ABA- and turgor-inducible promoters, and the like.

5 As the DNA sequence between the transcription initiation site and the start of the coding sequence, i.e., the untranslated leader sequence, can influence gene expression, one may also wish to employ a particular leader sequence. Any leader sequence available to one of skill in the art may be employed. Preferred leader sequences direct optimum levels of expression of the attached gene, for
10 example, by increasing or maintaining mRNA stability and/or by preventing inappropriate initiation of translation (Joshi, Nucl. Acid Res., 15, 6643 (1987)). The choice of such sequences is at the discretion of those of skill in the art. Sequences that are derived from genes that are highly expressed in dicots, and in soybean in particular, are contemplated.

15 In some cases, extremely high expression of anthranilate synthase or a domain thereof, is not necessary. For example, using the methods of the invention such high levels of anthranilate synthase may be generated that the availability of substrate, rather than enzyme, may limit the levels of tryptophan generated. In such cases, more moderate or regulated levels of expression can be
20 selected by one of skill in the art. Such a skilled artisan can readily modulate or regulate the levels of expression, for example, by use of a weaker promoter or by use of a developmentally regulated or tissue specific promoter.

25 Nucleic acids encoding the anthranilate synthase of interest can also include a plastid transit peptide (e.g. SEQ ID NO:72 or 74) to facilitate transport of the anthranilate synthase polypeptide into plastids, for example, into chloroplasts. A nucleic acid encoding the selected plastid transit peptide (e.g. SEQ ID NO: 71 or 73) is generally linked in-frame with the coding sequence of the anthranilate synthase. However, the plastid transit peptide can be placed at either the N-terminal or C-terminal end of the anthranilate synthase.

30 Constructs also include the nucleic acid of interest (e.g. DNA encoding an anthranilate synthase) along with a nucleic acid sequence that acts as a transcription termination signal and that allows for the polyadenylation of the resultant mRNA. Such transcription termination signals are placed 3' or

downstream of the coding region of interest. Preferred transcription termination signals contemplated include the transcription termination signal from the nopaline synthase gene of *Agrobacterium tumefaciens* (Bevan et al., Nucl. Acid Res., **11**, 369 (1983)), the terminator from the octopine synthase gene of 5 *Agrobacterium tumefaciens*, and the 3' end of genes encoding protease inhibitor I or II from potato or tomato, although other transcription termination signals known to those of skill in the art are also contemplated. Regulatory elements such as Adh intron 1 (Callis et al., Genes Develop., **1**, 1183 (1987)), sucrose synthase intron (Vasil et al., Plant Physiol., **91**, 5175 (1989)) or TMV omega 10 element (Gallie et al., The Plant Cell, **1**, 301 (1989)) may further be included where desired. These 3' nontranslated regulatory sequences can be obtained as described in An, Methods in Enzymology, **153**, 292 (1987) or are already present in plasmids available from commercial sources such as Clontech, Palo Alto, California. The 3' nontranslated regulatory sequences can be operably linked to 15 the 3 terminus of an anthranilate synthase gene by standard methods. Other such regulatory elements useful in the practice of the invention are known to those of skill in the art.

Selectable marker genes or reporter genes are also useful in the present 20 invention. Such genes can impart a distinct phenotype to cells expressing the marker gene and thus allow such transformed cells to be distinguished from cells that do not have the marker. Selectable marker genes confer a trait that one can 'select' for by chemical means, i.e., through the use of a selective agent (e.g., a herbicide, antibiotic, or the like). Reporter genes, or screenable genes, confer a trait that one can identify through observation or testing, i.e., by 'screening' (e.g., 25 the R-locus trait). Of course, many examples of suitable marker genes are known to the art and can be employed in the practice of the invention.

Possible selectable markers for use in connection with the present 30 invention include, but are not limited to, a *neo* gene (Potrykus et al., Mol. Gen. Genet., **199**, 183 (1985)) which codes for neomycin resistance and can be selected for using kanamycin, G418, and the like; a *bar* gene which codes for bialaphos resistance; a gene which encodes an altered EPSP synthase protein (Hinchee et al., Biotech., **6**, 915 (1988)) thus conferring glyphosate resistance; a nitrilase gene such as *bxn* from *Klebsiella ozaenae* which confers resistance to

bromoxynil (Stalker et al., *Science*, **242**, 419 (1988)); a mutant acetolactate synthase gene (ALS) that confers resistance to imidazolinone, sulfonylurea or other ALS-inhibiting chemicals (European Patent Application 154,204, 1985); a methotrexate-resistant DHFR gene (Thillet et al., *J. Biol. Chem.*, **263**, 12500 (1988)); a dalapon dehalogenase gene that confers resistance to the herbicide dalapon; or a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan. Where a mutant EPSP synthase gene is employed, additional benefit may be realized through the incorporation of a suitable plastid transit peptide (CTP).

10 An illustrative embodiment of a selectable marker gene capable of being used in systems to select transformants is the genes that encode the enzyme phosphinothricin acetyltransferase, such as the *bar* gene from *Streptomyces hygroscopicus* or the *pat* gene from *Streptomyces viridochromogenes* (U.S. Pat. No. 5,550,318, which is incorporated by reference herein). The enzyme phosphinothricin acetyl transferase (PAT) inactivates the active ingredient in the herbicide bialaphos, phosphinothricin (PPT). PPT inhibits glutamine synthetase, (Murakami et al., *Mol. Gen. Genet.*, **205**, 42 (1986); Twell et al., *Plant Physiol.*, **91**, 1270 (1989)) causing rapid accumulation of ammonia and cell death.

Screenable markers that may be employed include, but are not limited to, 20 a β -glucuronidase or *uidA* gene (GUS) which encodes an enzyme for which various chromogenic substrates are known; an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta et al., in *Chromosome Structure and Function*, pp. 263-282 (1988)); a β -lactamase gene (Sutcliffe, *PNAS USA*, **75**, 3737 (1978)), which 25 encodes an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin); a *xy/E* gene (Zukowsky et al., *PNAS USA*, **80**, 1101 (1983)) that encodes a catechol dioxygenase that can convert chromogenic catechols; an α -amylase gene (Ikuta et al., *Biotech.*, **8**, 241 (1990)); a tyrosinase gene (Katz et al., *J. Gen. Microbiol.*, **129**, 2703 (1983)) that encodes 30 an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily detectable compound melanin; a β -galactosidase gene, which encodes an enzyme for which there are chromogenic substrates; a luciferase (*lux*) gene (Ow et al., *Science*, **234**, 856 (1986)), which

allows for bioluminescence detection; or even an aequorin gene (Prasher et al., *Biochem. Biophys. Res. Comm.*, 126, 1259 (1985)), which may be employed in calcium-sensitive bioluminescence detection, or a green fluorescent protein gene (Niedz et al., *Plant Cell Reports*, 14, 403 (1995)). The presence of the *lux* gene 5 in transformed cells may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon-counting cameras, or multiwell luminometry. It is also envisioned that this system may be developed for populational screening for bioluminescence, such as on tissue culture plates, or even for whole plant screening.

10 Additionally, transgenes may be constructed and employed to provide targeting of the gene product to an intracellular compartment within plant cells or in directing a protein to the extracellular environment. This will generally be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of a particular gene. The resultant transit, or 15 signal, peptide will transport the protein to a particular intracellular, or extracellular destination, respectively, and may then be post-translationally removed. Transit or signal peptides act by facilitating the transport of proteins through intracellular membranes, e.g., vacuole, vesicle, plastid and mitochondrial membranes, whereas signal peptides direct proteins through the 20 extracellular membrane. By facilitating transport of the protein into compartments inside or outside the cell, these sequences may increase the accumulation of gene product.

25 A particular example of such a use concerns the direction of an anthranilate synthase to a particular organelle, such as the plastid, rather than to the cytoplasm. This is exemplified by the use of the *Arabidopsis* SSU1A transit peptide that confers plastid-specific targeting of proteins. Alternatively, the transgene can comprise a plastid transit peptide-encoding DNA sequence or a 30 DNA sequence encoding the *rbcS* (RuBISCO) transit peptide operably linked between a promoter and the DNA sequence encoding an anthranilate synthase (for a review of plastid targeting peptides, see Heijne et al., *Eur. J. Biochem.*, 180, 535 (1989); Keegstra et al., *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 40, 471 (1989)). If the transgene is to be introduced into a plant cell, the transgene can also contain plant transcriptional termination and polyadenylation signals

and translational signals linked to the 3' terminus of a plant anthranilate synthase gene.

An exogenous plastid transit peptide can be used which is not encoded within a native plant anthranilate synthase gene. A plastid transit peptide is 5 typically 40 to 70 amino acids in length and functions post-translationally to direct a protein to the plastid. The transit peptide is cleaved either during or just after import into the plastid to yield the mature protein. The complete copy of a gene encoding a plant anthranilate synthase may contain a plastid transit peptide sequence. In that case, it may not be necessary to combine an exogenously 10 obtained plastid transit peptide sequence into the transgene.

Exogenous plastid transit peptide encoding sequences can be obtained from a variety of plant nuclear genes, so long as the products of the genes are expressed as preproteins comprising an amino terminal transit peptide and transported into plastid. Examples of plant gene products known to include such 15 transit peptide sequences include, but are not limited to, the small subunit of ribulose biphosphate carboxylase, chlorophyll a/b binding protein, plastid ribosomal proteins encoded by nuclear genes, certain heatshock proteins, amino acid biosynthetic enzymes such as acetolactate acid synthase, 3-enolpyruvylphosphoshikimate synthase, dihydrodipicolinate synthase, 20 anthranilate synthase and the like. In some instances a plastid transport protein already may be encoded in the anthranilate synthase gene of interest, in which case there may be no need to add such plastid transit sequences. Alternatively, the DNA fragment coding for the transit peptide may be chemically synthesized either wholly or in part from the known sequences of transit peptides such as 25 those listed above.

Regardless of the source of the DNA fragment coding for the transit peptide, it should include a translation initiation codon, for example, an ATG codon, and be expressed as an amino acid sequence that is recognized by and will function properly in plastids of the host plant. Attention should also be 30 given to the amino acid sequence at the junction between the transit peptide and the anthranilate synthase enzyme where it is cleaved to yield the mature enzyme. Certain conserved amino acid sequences have been identified and may serve as a guideline. Precise fusion of the transit peptide coding sequence with the

anthranilate synthase coding sequence may require manipulation of one or both DNA sequences to introduce, for example, a convenient restriction site. This may be accomplished by methods including site-directed mutagenesis, insertion of chemically synthesized oligonucleotide linkers, and the like.

5 Precise fusion of the nucleic acids encoding the plastid transport protein may not be necessary so long as the coding sequence of the plastid transport protein is in-frame with that of the anthranilate synthase. For example, additional peptidyl or amino acids can often be included without adversely affecting the expression or localization of the protein of interest.

10 Once obtained, the plastid transit peptide sequence can be appropriately linked to the promoter and an anthranilate synthase coding region in a transgene using standard methods. A plasmid containing a promoter functional in plant cells and having multiple cloning sites downstream can be constructed or obtained from commercial sources. The plastid transit peptide sequence can be 15 inserted downstream from the promoter using restriction enzymes. An anthranilate synthase coding region can then be translationally fused or inserted immediately downstream from and in frame with the 3' terminus of the plastid transit peptide sequence. Hence, the plastid transit peptide is preferably linked to the amino terminus of the anthranilate synthase. Once formed, the transgene can 20 be subcloned into other plasmids or vectors.

In addition to nuclear plant transformation, the present invention also extends to direct transformation of the plastid genome of plants. Hence, targeting of the gene product to an intracellular compartment within plant cells may also be achieved by direct delivery of a gene to the intracellular compartment. Direct 25 transformation of plastid genome may provide additional benefits over nuclear transformation. For example, direct plastid transformation of anthranilate synthase eliminates the requirement for a plastid targeting peptide and post-translational transport and processing of the pre-protein derived from the corresponding nuclear transformants. Plastid transformation of plants has been 30 described by P. Maliga. Current Opinion in Plant Biology 5, 164-172 (2002), P. B. Heifetz. Biochimie vol. 82, 655-666 (2000), R. Bock. J. Mol. Biol. 312, 425-438 (2001), and H. Daniell et al., Trends in Plant Science 7, 84-91 (2002) and references within.

After constructing a transgene containing an anthranilate synthase gene, the cassette can then be introduced into a plant cell. Depending on the type of plant cell, the level of gene expression, and the activity of the enzyme encoded by the gene, introduction of DNA encoding an anthranilate synthase into the 5 plant cell can lead to the overproduction of tryptophan, confer tolerance to an amino acid analog of tryptophan, such as 5-methyltryptophan or 6-methylanthranilate, and/or otherwise alter the tryptophan content of the plant cell.

10 Transformation of Host Cells

A transgene comprising an anthranilate synthase gene can be subcloned into a known expression vector, and AS expression can be detected and/or quantitated. This method of screening is useful to identify transgenes providing for an expression of an anthranilate synthase gene, and expression of an 15 anthranilate synthase in the plastid of a transformed plant cell.

Plasmid vectors include additional DNA sequences that provide for easy selection, amplification, and transformation of the transgene in prokaryotic and eukaryotic cells, e.g., pUC-derived vectors, pSK-derived vectors, pGEM-derived vectors, pSP-derived vectors, or pBS-derived vectors. The additional DNA 20 sequences include origins of replication to provide for autonomous replication of the vector, selectable marker genes, preferably encoding antibiotic or herbicide resistance, unique multiple cloning sites providing for multiple sites to insert DNA sequences or genes encoded in the transgene, and sequences that enhance transformation of prokaryotic and eukaryotic cells.

Another vector that is useful for expression in both plant and prokaryotic 25 cells is the binary Ti plasmid (as disclosed in Schilperoort et al., U.S. Patent No. 4,940,838, issued July 10, 1990) as exemplified by vector pGA582. This binary Ti plasmid vector has been previously characterized by An, cited *supra*. This binary Ti vector can be replicated in prokaryotic bacteria such as *E. coli* and 30 *Agrobacterium*. The *Agrobacterium* plasmid vectors can also be used to transfer the transgene to plant cells. The binary Ti vectors preferably include the nopaline T DNA right and left borders to provide for efficient plant cell transformation, a selectable marker gene, unique multiple cloning sites in the T

border regions, the *colE1* replication of origin and a wide host range replicon.

The binary Ti vectors carrying a transgene of the invention can be used to transform both prokaryotic and eukaryotic cells, but is preferably used to transform plant cells. See, for example, Glassman et al., U.S. Pat. No.

5 5,258,300.

The expression vector can then be introduced into prokaryotic or eukaryotic cells by available methods. Methods of transformation especially effective for monocots and dicots, include, but are not limited to, microprojectile bombardment of immature embryos (U.S. Pat. No. 5,990,390) or Type II

- 10 embryogenic callus cells as described by W.J. Gordon-Kamm et al. (Plant Cell, 2, 603 (1990)), M.E. Fromm et al. (Bio/Technology, 8, 833 (1990)) and D.A. Walters et al. (Plant Molecular Biology, 18, 189 (1992)), or by electroporation of type I embryogenic calluses described by D'Halluin et al. (The Plant Cell, 4, 1495 (1992)), or by Krzyzek (U.S. Patent No. 5,384,253, issued January 24,
- 15 1995). Transformation of plant cells by vortexing with DNA-coated tungsten whiskers (Coffey et al., U.S. Patent No. 5,302,523, issued April 12, 1994) and transformation by exposure of cells to DNA-containing liposomes can also be used.

After transformation of the selected anthranilate synthase construct into a host cell, the host cell may be used for production of useful products generated by the transgenic anthranilate synthase in combination with the host cell's enzymatic machinery. Culturing the transformed cells can lead to enhanced production of tryptophan and other useful compounds, which can be recovered from the cells or from the culture media. Examples of useful compounds that may be generated upon expression in a variety of host cells and/or organisms include tryptophan, indole acetic acid and other auxins, isoflavanoid compounds important to cardiovascular health found in soy, volatile indole compounds which act as signals to natural enemies of herbivorous insects in maize, anticarcinogens such as indole glucosinolates (indole-3-carbinol) found in the Cruciferae plant family, as well as indole alkaloids such as ergot compounds produced by certain species of fungi. (Barnes et al., Adv Exp Med Biol, 401, 87 (1996); Frey et al., Proc Natl Acad Sci, 97, 14801 (2000); Muller et al., Biochem, 381, 679 (2000); Mantegani et al., Farmaco, 54, 288 (1999); Zeligs, J

Med Food, 1, 67 (1998); Mash et al., Ann NY Acad Sci, 844, 274 (1998);
Melanson et al., Proc Natl Acad Sci, 94, 13345 (1997); Broadbent et al., Curr
Med Chem, 5, 469 (1998).

Accumulation of tryptophan may also lead to the increased production of
5 secondary metabolites in microbes and plants, for example, indole containing
metabolites such as simple indoles, indole conjugates, indole alkaloids, indole
phytoalexins and indole glucosinolates in plants.

Anthraniilate synthases insensitive to tryptophan have the potential to
increase a variety of chorismate-derived metabolites, including those derived
10 from phenylalanine due to the stimulation of phenylalanine synthesis by
tryptophan via chorismate mutase. *See Siehl, D. The biosynthesis of tryptophan,
tyrosine, and phenylalanine from chorismate in Plant Amino Acids:
Biochemistry and Biotechnology*, ed. BK Singh, pp 171-204. Other chorismate-
derived metabolites that may increase when feedback insensitive anthraniilate
15 synthases are present include phenylpropanoids, flavonoids, and isoflavonoids,
as well as those derived from anthraniilate, such as indole, indole alkaloids, and
indole glucosinolates. Many of these compounds are important plant hormones,
plant defense compounds, chemopreventive agents of various health conditions,
and/or pharmacologically active compounds.

20 The range of these compounds whose synthesis might be increased by
expression of anthraniilate synthase depends on the organism in which the
anthraniilate synthase is expressed. One of skill in the art can readily assess
which organisms and host cells to use and/or test in order to generate the desired
compounds. The invention contemplates synthesis of tryptophan and other
25 useful compounds in a variety of organisms, including plants, microbes, fungi,
yeast, bacteria, insect cells, and mammalian cells.

Strategy for Selection of Tryptophan Overproducer Cell Lines

Efficient selection of a desired tryptophan analog resistant, tryptophan
30 overproducer variant using tissue culture techniques requires careful
determination of selection conditions. These conditions are optimized to allow
growth and accumulation of tryptophan analog resistant, tryptophan
overproducer cells in the culture while inhibiting the growth of the bulk of the

cell population. The situation is complicated by the fact that the vitality of individual cells in a population can be highly dependent on the vitality of neighboring cells.

Conditions under which cell cultures are exposed to tryptophan analog
5 are determined by the characteristics of the interaction of the compound with the tissue. Such factors as the degree of toxicity and the rate of inhibition should be considered. The accumulation of the compounds by cells in culture, and the persistence and stability of the compounds, both in the media and in the cells, also need to be considered along with the extent of uptake and transmission to
10 the desired cellular compartment. Additionally, it is important to determine whether the effects of the compounds can be readily reversed by the addition of tryptophan.

The effects of the analog on culture viability and morphology is carefully evaluated. It is especially important to choose analog exposure conditions that
15 have no impact on plant regeneration capability of cultures. Choice of analog exposure conditions is also influenced by whether the analog kills cells or simply inhibits cell divisions.

The choice of a selection protocol is dependent upon the considerations described above. The protocols briefly described below can be utilized in the
20 selection procedure. For example, to select for cells that are resistant to growth inhibition by a tryptophan analog, finely divided cells in liquid suspension culture can be exposed to high tryptophan analog levels for brief periods of time. Surviving cells are then allowed to recover and accumulate and are then reexposed for subsequently longer periods of time. Alternatively, organized
25 partially differentiated cell cultures are grown and subcultured with continuous exposure to initially low levels of a tryptophan analog. Concentrations are then gradually increased over several subculture intervals. While these protocols can be utilized in a selection procedure, the present invention is not limited to these procedures.

30

Genes for Plant Modification

As described hereinabove, genes that function as selectable marker genes and reporter genes can be operably combined with the DNA sequence encoding

the anthranilate synthase, or domain thereof, in transgenes, vectors and plants of the present invention. Additionally, other agronomical traits can be added to the transgenes, vectors and plants of the present invention. Such traits include, but are not limited to, insect resistance or tolerance; disease resistance or tolerance (viral, bacterial, fungal, nematode); stress resistance or tolerance, as exemplified by resistance or tolerance to drought, heat, chilling, freezing, excessive moisture, salt stress, oxidative stress; increased yields; food content and makeup; physical appearance; male sterility; drydown; standability; prolificacy; starch properties; oil quantity and quality; and the like. One may incorporate one or more genes conferring such traits into the plants of the invention.

Insect Resistance or Tolerance

Bacillus thuringiensis (or "Bt") bacteria include nearly 20 known subspecies of bacteria which produce endotoxin polypeptides that are toxic when ingested by a wide variety of insect species. The biology and molecular biology of the endotoxin proteins (Bt proteins) and corresponding genes (Bt genes) has been reviewed by H. R. Whately et al., *Ann. Rev. Microbiol.*, **40**, 549 (1986) and by H. Hofte et al., *Microbiol. Rev.*, **53**, 242 (1989). Genes coding for a variety of Bt proteins have been cloned and sequenced. A segment of the Bt polypeptide is essential for toxicity to a variety of *Lepidoptera* pests and is contained within approximately the first 50% of the Bt polypeptide molecule. Consequently, a truncated Bt polypeptide coded by a truncated Bt gene will in many cases retain its toxicity towards a number of *Lepidoptera* insect pests. For example, the HD73 and HD1 Bt polypeptides have been shown to be toxic to the larvae of the important *Lepidoptera* insect pests of plants in the USA such as the European corn borer, cutworms and earworms. The genes coding for the HD1 and HD73 Bt polypeptides have been cloned and sequenced by M. Geiser et al., *Gene*, **48**, 109 (1986) and M. J. Adang et al., *Gene*, **36**, 289 (1985), respectively, and can be cloned from HD1 and HD73 strains obtained from culture collections (e.g. *Bacillus* Genetic Stock Center, Columbus, Ohio or USDA Bt stock collection Peoria, Ill.) using standard protocols. Examples of Bt genes and polypeptides are described, for example, in U.S. Patent Numbers 6,329,574, 6,303,364, 6,320,100 and 6,331,655.

DNA coding for new, previously uncharacterized Bt toxins, may be cloned from the host *Bacillus* organism using protocols that have previously been used to clone Bt genes, and new synthetic forms of Bt toxins may also be produced.

- 5 A Bt gene useful in the present invention may include a 5' DNA sequence including a sequence of DNA which will allow for the initiation of transcription and translation of a downstream located Bt sequence in a plant. The Bt gene may also comprise a 3' DNA sequence that includes a sequence derived from the 3' non-coding region of a gene that can be expressed in the plant of interest. The Bt
10 gene would also include a DNA sequence coding for a toxic Bt polypeptide produced by *Bacillus thuringiensis* or toxic portions thereof or having substantial amino sequence homology thereto. The Bt coding sequence may include: (i) DNA sequences which code for insecticidal proteins that have substantial homology to Bt endotoxins that are active against insect pests of the plant of
15 interest, e.g., the HD73 or HD1 Bt sequences; (ii) sequences coding for insecticidally-active segments of the Bt endotoxin polypeptide, e.g., insecticidally active HD73 or HD1 polypeptides truncated from the carboxy and/or amino termini; and/or (iii) a truncated Bt sequence fused in frame with a sequence(s) that codes for a polypeptide that provides some additional advantage
20 such as: (a) genes that are selectable, e.g., genes that confer resistance to antibiotics or herbicides, (b) reporter genes whose products are easy to detect or assay, e.g., luciferase or beta-glucuronidase; (c) DNA sequences that code for polypeptide sequences that have some additional use in stabilizing the Bt protein against degradation or enhance the efficacy of the Bt protein against insects, e.g.,
25 protease inhibitors and (d) sequences that help direct the Bt protein to a specific compartment inside or outside the plant cell, e.g., a signal sequence.

To obtain optimum synthesis of the Bt protein in the plant, it may also be appropriate to adjust the DNA sequence of the Bt gene to more resemble the genes that are efficiently expressed in the plant of interest. Since the codon usage
30 of Bt genes may be dissimilar to that used by genes that are expressed in the plant of interest, the expression of the Bt gene in plant cells may be improved by the replacement of these codons with those that are more efficiently expressed in plants, e.g., are used more frequently in the plants of interest (See E. Murray et

al., Nucl. Acids Res., **17**, 477 (1989)). Such replacement of codons may require the substitution of bases without changing the amino acid sequence of the resulting Bt polypeptide. The Bt polypeptide may be identical in sequence to the bacterial gene or segments thereof. The complete Bt coding sequence, or sections 5 thereof, containing a higher proportion of preferred codons than the original bacterial gene could be synthesized using standard chemical synthesis protocols, and introduced or assembled into the Bt gene using standard protocols, such as site-directed mutagenesis or DNA polymerization and ligation and the like.

Protease inhibitors may also provide insect resistance. For example, use 10 of a protease inhibitor II gene, pinII, from tomato or potato may be useful. Also advantageous is the use of a pinII gene in combination with a Bt toxin gene. Other genes which encode inhibitors of the insects' digestive system, or those that encode enzymes or co-factors that facilitate the production of inhibitors, may also be useful. This group includes oryzacystatin and amylase inhibitors such as 15 those from wheat and barley.

Genes encoding lectins may confer additional or alternative insecticide properties. (Murdock et al., Phytochemistry, **29**, 85 (1990); Czapla & Lang, I. Econ. Entomol., **83**, 2480 (1990) Lectin genes contemplated to be useful include, for example, barley and wheat germ agglutinin (WGA) and rice lectins. 20 (Gatehouse et al., J Sci Food Agric., **35**, 373 (1984))

Genes controlling the production of large or small polypeptides active against insects when introduced into the insect pests such as lytic peptides, peptide hormones and toxins and venoms, may also be useful. For example, the expression of juvenile hormone esterase, directed towards specific insect pests, 25 may also result in insecticidal activity, or perhaps cause cessation of metamorphosis. (Hammock et al., Nature, **344**, 458 (1990))

Transgenic plants expressing genes encoding enzymes that affect the 30 integrity of the insect cuticle may also be useful. Such genes include those encoding, for example, chitinase, proteases, lipases and also genes for the production of nikkomycin. Genes that code for activities that affect insect molting, such those affecting the production of ecdysteroid UDP-glucosyl transferase, may also be useful.

Genes that code for enzymes that facilitate the production of compounds that reduce the nutritional quality of the plant to insect pests a may also be useful. It may be possible, for instance, to confer insecticidal activity to a plant by altering its sterol composition. Further embodiments of the invention concern 5 transgenic plants with enhanced lipoxygenase activity.

The present invention also provides methods and compositions useful in altering plant secondary metabolites. One example concerns altering plants to produce DIMBOA which, it is contemplated, will confer resistance to European corn borer, rootworm and several other insect pests. See, e.g., U.S. Patent 10 6,331,880. DIMBOA is derived from indole-related compounds. The present invention provides methods for increasing the content of indole-related compounds like tryptophan within plant cells and tissues. Hence, according to the invention the methods provided herein may also increase the levels of DIMBOA, and thereby increase the resistance of plants to insects.

15 The introduction of genes that can regulate the production of maysin, and genes involved in the production of dhurrin in sorghum, is also contemplated to be of use in facilitating resistance to earworm and rootworm, respectively.

Further genes encoding proteins characterized as having potential insecticidal activity may also be used. Such genes include, for example, the 20 cowpea trypsin inhibitor (CpTI; Hilder et al., *Nature*, 330, 160 (1987)) which may be used as a rootworm deterrent; genes encoding avermectin (Avermectin and Abamectin., Campbell, W. C., Ed., 1989; Ikeda et al., *J Bacteriol*, 169, 5615 1987) which may prove useful as a corn rootworm deterrent; ribosome 25 inactivating protein genes; and genes that regulate plant structures. Transgenic plants including anti-insect antibody genes and genes that code for enzymes that can convert a non-toxic insecticide (pro-insecticide) applied to the outside of the plant into an insecticide inside the plant are also contemplated.

Environmental or Stress Resistance or Tolerance

Improvement of a plant's ability to tolerate various environmental 30 stresses can be effected through expression of genes. For example, increased resistance to freezing temperatures may be conferred through the introduction of an "antifreeze" protein such as that of the Winter Flounder (Cutler et al., *J Plant Physiol*, 135, 351 1989) or synthetic gene derivatives thereof. Improved chilling

tolerance may also be conferred through increased expression of glycerol-3-phosphate acetyltransferase in plastids (Wolter et al., The EMBO J., **11**, 4685 (1992)). Resistance to oxidative stress can be conferred by expression of superoxide dismutase (Gupta et al., Proc. Natl. Acad. Sci USA, **90**, 1629 (1993)),
5 and can be improved by glutathione reductase (Bowler et al., Ann Rev. Plant Physiol., **43**, 83 (1992)).

It is contemplated that the expression of genes that favorably affect plant water content, total water potential, osmotic potential, and turgor will enhance the ability of the plant to tolerate drought and will therefore be useful. It is
10 proposed, for example, that the expression of genes encoding for the biosynthesis of osmotically-active solutes may impart protection against drought. Within this class are genes encoding for mannitol dehydrogenase (Lee and Saier, J. Bacteriol., **258**, 10761 (1982)) and trehalose-6-phosphate synthase (Kaasen et al., J. Bacteriology, **174**, 889 (1992)).

15 Similarly, other metabolites may protect either enzyme function or membrane integrity (Loomis et al., J. Expt. Zoology, **252**, 9 (1989)), and therefore expression of genes encoding for the biosynthesis of these compounds might confer drought resistance in a manner similar to or complimentary to mannitol. Other examples of naturally occurring metabolites that are osmotically
20 active and/or provide some direct protective effect during drought and/or desiccation include fructose, erythritol, sorbitol, dulcitol, glucosylglycerol, sucrose, stachyose, raffinose, proline, glycine, betaine, ononitol and pinitol. See, e.g., U.S. Patent 6,281,411.

Three classes of Late Embryogenic Proteins have been assigned based on
25 structural similarities (see Dure et al., Plant Molecular Biology, **12**, 475 (1989)). Expression of structural genes from all three LEA groups may confer drought tolerance. Other types of proteins induced during water stress, which may be useful, include thiol proteases, aldolases and transmembrane transporters, which may confer various protective and/or repair-type functions during drought stress.
30 See, e.g., PCT/CA99/00219 (Na⁺/H⁺ exchanger polypeptide genes). Genes that effect lipid biosynthesis might also be useful in conferring drought resistance.

The expression of genes involved with specific morphological traits that allow for increased water extractions from drying soil may also be useful. The

expression of genes that enhance reproductive fitness during times of stress may also be useful. It is also proposed that expression of genes that minimize kernel abortion during times of stress would increase the amount of grain to be harvested and hence be of value.

5 Enabling plants to utilize water more efficiently, through the introduction and expression of genes, may improve the overall performance even when soil water availability is not limiting. By introducing genes that improve the ability of plants to maximize water usage across a full range of stresses relating to water availability, yield stability or consistency of yield performance may be realized.

10 **Disease Resistance or Tolerance**

Resistance to viruses may be produced through expression of genes. For example, expression of antisense genes targeted at essential viral functions or expression of genes encoding viral coat proteins may impart resistance to the virus.

15 Resistance to diseases caused by bacteria and fungi may be conferred through introduction of genes. For example, genes encoding so-called "peptide antibiotics," pathogenesis related (PR) proteins, toxin resistance, and proteins affecting host-pathogen interactions such as morphological characteristics may be useful.

20 **Mycotoxin Reduction/Elimination**

Production of mycotoxins, including aflatoxin and fumonisins, by fungi associated with plants is a significant factor in rendering grain not useful.

Inhibition of the growth of these fungi may reduce the synthesis of these toxic substances and therefore reduce grain losses due to mycotoxin contamination. It 25 may be possible to introduce genes into plants such that would inhibit synthesis of the mycotoxin without interfering with fungal growth. Further, expression of a novel gene which encodes an enzyme capable of rendering the mycotoxin nontoxic would be useful in order to achieve reduced mycotoxin contamination of grain.

30 **Plant Composition or Quality**

The composition of the plant may be altered, for example, to improve the balance of amino acids in a variety of ways including elevating expression of native proteins, decreasing expression of those with poor composition, changing

- the composition of native proteins, or introducing genes encoding entirely new proteins possessing superior composition. See, e.g., U.S. Patent No. 6,160,208 (alteration of seed storage protein expression). The introduction of genes that alter the oil content of the plant may be of value. See, e.g., U.S. Patent Nos. 5 6,069,289 and 6,268,550 (ACCase gene). Genes may be introduced that enhance the nutritive value of the starch component of the plant, for example by increasing the degree of branching, resulting in improved utilization of the starch in cows by delaying its metabolism.

Plant Agronomic Characteristics

- 10 Two of the factors determining where plants can be grown are the average daily temperature during the growing season and the length of time between frosts. Expression of genes that are involved in regulation of plant development may be useful, e.g., the liguleless and rough sheath genes that have been identified in corn.
- 15 Genes may be introduced into corn that would improve standability and other plant growth characteristics. Expression of genes which confer stronger stalks, improved root systems, or prevent or reduce ear dropage would be of value to the farmer

Nutrient Utilization

- 20 The ability to utilize available nutrients may be a limiting factor in growth of plants. It may be possible to alter nutrient uptake, tolerate pH extremes, mobilization through the plant, storage pools, and availability for metabolic activities by the introduction of genes. These modifications would allow a plant to more efficiently utilize available nutrients. For example, an 25 increase in the activity of an enzyme that is normally present in the plant and involved in nutrient utilization may increase the availability of a nutrient. An example of such an enzyme would be phytase.

Male Sterility

- Male sterility is useful in the production of hybrid seed, and male sterility 30 may be produced through expression of genes. It may be possible through the introduction of TURF-13 via transformation to separate male sterility from disease sensitivity. See Levings, Science, 250:942-947, 1990. As it may be

necessary to restore male fertility for breeding purposes and for grain production, genes encoding restoration of male fertility may also be introduced.

Selection and Characterization of Resistant Cell Lines

- 5 Selections are carried out until cells or tissue are recovered which are observed to be growing well in the presence of normally inhibitory levels of a tryptophan analog thereof. These cell "lines" are subcultured several additional times in the presence of a tryptophan analog to remove non-resistant cells and then characterized. The amount of resistance that has been obtained is
- 10 determined by comparing the growth of these cell lines with the growth of unselected cells or tissue in the presence of various tryptophan analogs at various concentrations. Stability of the resistance trait of the cultured cells may be evaluated by simply growing the selected cell lines in the absence of the tryptophan analog for various periods of time and then analyzing growth after re-
- 15 exposing the tissue to the analog. The resistant cell lines may also be evaluated using *in vitro* chemical studies to verify that the site of action of the analog is altered to a form that is less sensitive to inhibition by tryptophan analogs.

- Transient expression of an anthranilate synthase gene can be detected and quantitated in the transformed cells. Gene expression can be quantitated by RT-PCR analysis, a quantitative Western blot using antibodies specific for the cloned anthranilate synthase or by detecting enzyme activity in the presence of tryptophan or an amino acid analog of tryptophan. The tissue and subcellular location of the cloned anthranilate synthase can be determined by immunochemical staining methods using antibodies specific for the cloned anthranilate synthase or subcellular fractionation and subsequent biochemical and/or immunological analyses. Sensitivity of the cloned anthranilate synthase to agents can also be assessed. Transgenes providing for expression of an anthranilate synthase or anthranilate synthase tolerant to inhibition by an amino acid analog of tryptophan or free L-tryptophan can then be used to transform 20 monocot and/or dicot plant tissue cells and to regenerate transformed plants and seeds. Transformed cells can be selected by detecting the presence of a selectable marker gene or a reporter gene, for example, by detecting a selectable herbicide resistance marker. Transient expression of an anthranilate synthase
- 25
- 30

gene can be detected in the transgenic embryogenic calli using antibodies specific for the cloned anthranilate synthase, or by RT-PCR analyses.

Plant Regeneration and Production of Seed

- 5 Transformed embryogenic calli, meristematic tissue, embryos, leaf discs and the like can then be used to generate transgenic plants that exhibit stable inheritance of the transformed anthranilate synthase gene. Plant cell lines exhibiting satisfactory levels of tolerance to an amino acid analog of tryptophan are put through a plant regeneration protocol to obtain mature plants and seeds
- 10 expressing the tolerance traits by methods well known in the art (for example, see U.S. Pat. Nos. 5,990,390, 5,489,520; and Laursen et al., *Plant Mol. Biol.*, **24**, 51 (1994)). The plant regeneration protocol allows the development of somatic embryos and the subsequent growth of roots and shoots. To determine that the tolerance trait is expressed in differentiated organs of the plant, and not solely in
- 15 undifferentiated cell culture, regenerated plants can be assayed for the levels of tryptophan present in various portions of the plant relative to regenerated, non-transformed plants. Transgenic plants and seeds can be generated from transformed cells and tissues showing a change in tryptophan content or in resistance to a tryptophan analog using standard methods. It is especially
- 20 preferred that the tryptophan content of the leaves or seeds is increased. A change in specific activity of the enzyme in the presence of inhibitory amounts of tryptophan or an analog thereof can be detected by measuring enzyme activity in the transformed cells as described by Widholm, *Biochimica et Biophysica Acta*, **279**, 48 (1972). A change in total tryptophan content can also be examined by
- 25 standard methods as described by Jones et al., *Analyst*, **106**, 968 (1981).
- Mature plants are then obtained from cell lines that are known to express the trait. If possible, the regenerated plants are self pollinated. In addition, pollen obtained from the regenerated plants is crossed to seed grown plants of agronomically important inbred lines. In some cases, pollen from plants of these
- 30 inbred lines is used to pollinate regenerated plants. The trait is genetically characterized by evaluating the segregation of the trait in first and later generation progeny. The heritability and expression in plants of traits selected in

tissue culture are of particular importance if the traits are to be commercially useful.

The commercial value of tryptophan overproducer soybeans, cereals and other plants is greatest if many different hybrid combinations are available for sale. The farmer typically grows more than one kind of hybrid based on such differences as maturity, standability or other agronomic traits. Additionally, hybrids adapted to one part of the country are not adapted to another part because of differences in such traits as maturity, disease, and insect resistance. Because of this, it is necessary to breed tryptophan overproduction into a large number of parental inbred lines so that many hybrid combinations can be produced.

A conversion process (backcrossing) is carried out by crossing the original overproducer line to normal elite lines and crossing the progeny back to the normal parent. The progeny from this cross will segregate such that some plants carry the gene responsible for overproduction whereas some do not. Plants carrying such genes will be crossed again to the normal parent resulting in progeny which segregate for overproduction and normal production once more. This is repeated until the original normal parent has been converted to an overproducing line, yet possesses all other important attributes as originally found in the normal parent. A separate backcrossing program is implemented for every elite line that is to be converted to tryptophan overproducer line.

Subsequent to the backcrossing, the new overproducer lines and the appropriate combinations of lines which make good commercial hybrids are evaluated for overproduction as well as a battery of important agronomic traits. Overproducer lines and hybrids are produced which are true to type of the original normal lines and hybrids. This requires evaluation under a range of environmental conditions where the lines or hybrids will generally be grown commercially. For production of high tryptophan soybeans, it may be necessary that both parents of the hybrid seed be homozygous for the high tryptophan character. Parental lines of hybrids that perform satisfactorily are increased and used for hybrid production using standard hybrid seed production practices.

The transgenic plants produced herein are expected to be useful for a variety of commercial and research purposes. Transgenic plants can be created for use in traditional agriculture to possess traits beneficial to the consumer of

the grain harvested from the plant (e.g., improved nutritive content in human food or animal feed). In such uses, the plants are generally grown for the use of their grain in human or animal foods. However, other parts of the plants, including stalks, husks, vegetative parts, and the like, may also have utility, 5 including use as part of animal silage, fermentation feed, biocatalysis, or for ornamental purposes.

Transgenic plants may also find use in the commercial manufacture of proteins or other molecules, where the molecule of interest is extracted or purified from plant parts, seeds, and the like. Cells or tissue from the plants may 10 also be cultured, grown *in vitro*, or fermented to manufacture such molecules.

The transgenic plants may also be used in commercial breeding programs, or may be crossed or bred to plants of related crop species. Improvements encoded by the recombinant DNA may be transferred, e.g., from soybean cells to cells of other species, e.g., by protoplast fusion.

15 In one embodiment, a transgene comprised of a maize anthranilate α -domain isolated from a maize cell line tolerant to 5-MT and linked to the 35S CaMV promoter is introduced into a 5-MT sensitive monocot or dicot tissue using microprojectile bombardment. Transformed embryos or meristems are selected and used to generate transgenic plants. Transformed calli and transgenic 20 plants can be evaluated for tolerance to 5-MT or 6-MA and for stable inheritance of the tolerance trait.

The following examples further illustrate the invention and are not intended to be limiting thereof.

25

EXAMPLE 1: Isolation and *E. coli* Expression of Anthranilate Synthase from *Agrobacterium tumefaciens*.

This example describes the isolation of anthranilate synthase from *Agrobacterium tumefaciens* and its expression in *E. coli*.

30

Cloning of *Agrobacterium tumefaciens* AS

The nucleotide and amino acid sequences of the anthranilate synthase coding region from *Rhizobium meliloti* (GenBank accession number: P15395)

was used to search an *Agrobacterium tumefaciens* C58 genomic sequence database (Goodner et al. *Science* 294, 2323-2328 (2001)). The search consisted of tblastn using blosum62 matrix, (Altschul et. al., *Nucleic Acid Res.*, 25, 3389-3402 (1997)).

5 The identified AS homolog in the *Agrobacterium tumefaciens* C58 genomic sequence database was cloned by PCR using genomic DNA from *Agrobacterium tumefaciens* strain C58 (ATCC No. 33970) as the template. The primary PCR reaction was carried out using the following primers:

5'-TTATGCCGCCTGTCATCG-3' (SEQ ID NO:47) and

10 5'-ATAGGCTTAATGGTAACCG-3' (SEQ ID NO:48).

Gene amplification parameters were as follows: (a) denature at 95°C for 30 seconds, (b) anneal at 50°C for 30 seconds and (c) extend at 72 °C for 2 minutes, using Expand high fidelity PCR (Roche Biochemicals), according to manufacturer directions.

15 An additional round of PCR amplification, yielding a product of approximately 2.3 Kb in length, was carried out using the amplified template from above and the following nested primers:

5'-CTGAAACAACAGAAAGTACG-3' (SEQ ID NO:49)

5'-TAACCGTGTTCATCGAGCG-3' (SEQ ID NO:50).

20 The purified PCR product was ligated into pGEM-T easy (Promega Biotech) resulting in the plasmid pMON61600 (Figure 1). pMON61600 was sequenced using standard sequencing methodology. Confirmation of the correct sequence was obtained by comparison of the sequence the *Rhizobium meliloti* anthranilate synthase sequence (Figure 2). The translated amino acid sequence 25 from the isolated clone (SEQ ID NO:4) shared 88% identity with the *Rhizobium meliloti* enzyme (SEQ ID NO:7) (Figure 2).

The abbreviation "AgroAS" or *A. tumefaciens* AS is sometimes used herein to refer to *Agrobacterium tumefaciens* anthranilate synthase.

30 **E. coli expression of Agrobacterium tumefaciens AS**

The following vectors were constructed to facilitate subcloning of the *Agrobacterium tumefaciens* AS gene into a suitable expression vector.

A 2215 base pair PCR fragment was generated using pMON61600 as the template and the following primers:

5'-AAAAAGATCTCCATGG TAACGATCATTCAAGG-3' (SEQ ID NO:51)

5'-AAAAGAA TTCTTATCACGCCCTGGTCTCGCC-3' (SEQ ID

5 NO:52).

The plasmid pMON61600 was digested with restriction enzymes NcoI and RsrII. In addition, a 409bp fragment (derived by digesting the 2215 base pair PCR product with NcoI and RsrII) was then ligated into the digested pMON61600 plasmid, thereby replacing the NcoI/RsrII fragment, and resulting 10 in a NcoI site in frame with the translation initiation codon (ATG) of *Agrobacterium tumefaciens* AS to yield plasmid pMON34692 (Figure 3).

15 The base T7 *E. coli* expression plasmid, pMON34697 (Figure 4), was generated by restriction digestion of pET30a (Novogen, Inc) with SphI and BamHI. The resulting 4,969 bp fragment was purified and subcloned with a 338 bp SphI and BamHI fragment from pET11d (Novogen, Inc).

The plasmid pMON34705 (Figure 5) was generated by restriction digestion of pMON34697 with NcoI and SacI. The resulting 5,263 bp fragment was then purified and ligated with a 2,256 bp NcoI and SacI fragment from pMON34692 containing *Agrobacterium tumefaciens* AS.

20 The plasmid pMON34705 was transformed into *E. coli* BL21(DE3) (*FompT HsdS_b(r_B m_B)gal dcm* (DE3)) according to manufacturer's instructions (Novogen, Inc). DE3 is a host lysogen of λDE3 containing chromosomal copy of T7 RNA polymerase under control of an isopropyl-1-thio-D-galactopyranoside (IPTG) inducible *lacUV5*.

25 Transformed cells were selected on kanamycin plates that had been incubated at 37°C overnight (10 hours). Single colonies were transferred to 2ml of LB (Luria Broth; per liter, 10g tryptone, 5g yeast extract, 10g NaCl, and 1g glucose (optional)) or 2X-YT broth (per liter, 16g tryptone, 10g yeast extract, 5g NaCl) and then placed in a 37°C incubator and shaken at 225rpm for 3 hours.

30 The cells were removed and 4μL of 100mM IPTG was added to the culture and returned to the 37°C incubator for an additional 2 to 3 hours. A 1mL aliquot of the cells was removed and sonicated in sonication buffer, (50mM potassium phosphate (pH 7.3), 10% glycerol, 10mM 2-mercaptoethanol and 10mM MgCl₂).

The resulting lysed cell extract was the source material for the standard AS assay described below. The results established that the expression system based on plasmid pMON34705 was able to produce soluble and enzymatically active *Agrobacterium tumefaciens* AS protein that accounts for approximately 50% of total soluble extracted protein.

**EXAMPLE 2: High Trp Seed Levels are Achieved by Transformation of
Plants with Wild Type Agrobacterium Anthranilate Synthase**

Expression Vector pMON58120

5 The vector pMON58120 (Figure 34) encodes a fusion between a 264 base pair *Arabidopsis* small subunit (SSU) chloroplast targeting peptide (CTP, SEQ ID NO:71) and a 2187 base pair wild type *Agrobacterium* anthranilate synthase (AgroAS) open reading frame (SEQ ID NO:1). See, Stark et al., (1992) Science 258: 287. Expression of this open reading frame is driven by the soy 7S 10 alpha prime (7S α) promoter.

Upon translation on cytoplasmic ribosomes, the fusion (immature protein) is imported into chloroplast where the chloroplast targeting sequence is removed. There are two cleavage sites in the CTP1. The first site is 30 base pairs upstream of the CDS start (C/M), and the other is at the initial methionine 15 (C/M). The second cleavage site does not seem to be processed efficiently. The cleavage is predicted to yield a mature protein of about 70Kd that has AS activity as shown by enzyme activity data and trp efficacy data.

The AS gene was transformed with the synthetic CP4 gene that confers 20 glyphosate resistance, however the CP4 gene is processed separately from the AS gene. Expression of the CP4 gene was driven by the FMV promoter, which is a 35S promoter from Figwort Mosaic Virus. Glyphosate resistance allows for selection of the transformed plants.

Western analysis of AS protein

25 Thirty-five transformation events of pMON58120 were analyzed for AgroAS protein presence. AgroAS protein was detected with a polyclonal antibody raised in rabbits against purified His-tagged AgroAS. The His-tagged, full-length Agro-AS polypeptide was used as an antigen to generate a population of polyclonal antibodies in rabbits by CoCalico Biological, INC. The 30 recombinant His-tagged Agro-AS DNA was placed into a pMON 34701 (pet-30a-agroAS) expression vector. The His-AgroAS fusion protein was expressed in *E.coli* BL21(DE3) and purified by Ni-NTA resin system (Qiagen protocol). For western analysis, primary rabbit anti-AgroAS antibodies were used at

1:5,000 dilution. Secondary, goat anti-rabbit alkaline phosphatase-conjugated antibodies were used at 1:5,000 dilution. In transgenic lines carrying 7 α lpha'-Agro AS genes, western blot analysis consistently revealed the presence of a single band that specifically cross-reacted with anti-AgroAS antibodies. This
5 band was not detected in the nontransgenic control line.

Free Amino Acid Analysis of Soy and *Arabidopsis* Seed

10 **Amino Acid Extraction:** About 50 mg of crushed soy seed (5 mg of *Arabidopsis*) material was placed in each centrifuge vial. One milliliter of 5% trichloroacetic acid was added to each sample (100 μ l for *Arabidopsis*). The samples were vortexed, and allowed to sit, with agitation, at room temperature for 15 min. They were then microcentrifuged for 15 min at 14000 rpm. Some of the supernate was then removed, placed in a HPLC vial and sealed. Samples were kept at 4°C in the analysis queue.

15 **Amino Acid Analysis:** The reagents utilized for amino acid analysis included the OPA reagent (o-phthalaldehyde and 3-mercaptopropionic acid in borate buffer (Hewlett-Packard, PN 5061-3335)) where the borate buffer (0.4 N in water, pH 10.2). The analysis was performed using the Agilent 1100 series HPLC system as described in the Agilent Technical Publication, "Amino Acid Analysis Using Zorbax Eclipse-AAA Columns and the Agilent 1100 HPLC." March 17, 2000. First, 0.5 μ l of the sample was derivatized with 2.5 μ l of OPA reagent in 10 μ l of borate buffer. Second, the derivative is injected onto a Eclipse XDB-C18 5 μ m, 4.6 x 150 mm column using a flow rate of 1.2 ml/min. Amino acid concentrations were measured using fluorescence: excitation at 340 nm, emission at 450 nm. Elution was with a gradient of HPLC Buffers A and B according to Table A, where HPLC Buffer A was 40 mM Na₂HPO₄, pH=7.8 and HPLC Buffer B was 9 : 9 : 2 :: Methanol : Acetonitrile : Water.

Table A: Amino Acid Elution

Time	0	20	21	26	27
% Buffer B	5	65	100	100	100

Amino acid standards were prepared from the dry chemicals, using all amino acids of interest. Proline analysis required an additional derivatization step with 9-fluorenylmethyl-chloroformate (FMOC). Amino acid standards were also sometimes purchased in concentrations ranging from 0 to 100 µg/ml. Samples 5 were reported in µg/g of seed powder. Calculations were performed using an MS Excel spreadsheet found on Mynabird TMBROW > Public > Calculators > External Standard.xls.

**Expression of Wild Type Agrobacterium Anthranilate Synthase in
10 Arabidopsis.**

The vector pMON 58120 was transformed into Arabidopsis plants by vacuum infiltration of the secondary inflorescences, and plants were allowed to set transgenic seed. The seed was collected and screened for the presence of a selectable marker (glyphosate resistance). Glyphosate resistant plants were 15 grown to maturity and seed from each plant, which was designated a transformation event, and analyzed for tryptophan content (Table B). Selected transformation events were also analyzed for the presence of the expressed *Agrobacterium* anthranilate synthase protein in the mature seed by Western blot analysis as shown in Table B.

20

Table B: Analysis of Transformants

Transformation Event	Trp (ppm)	Protein present
7317	2547	+
7315	2960	+
7319	3628	+
7313	3979	+

25 Expression of Wild Type Agrobacterium Anthranilate Synthase in Soy (Glycine Max)

Thirty-three out of thirty-five soy transformation events analyzed had an increase in seed trp levels, for example, from above 500 ppm and up to 12,000 ppm. In nontransgenic soy seeds, the trp level is less than 200 ppm. All seeds 30 that contained high amounts of trp demonstrated anthranilate synthase protein

expression by western blotting. Table C presents data for nineteen soy events that contain high trp levels and also are positive for anthranilate synthase anthranilate synthase protein by western blot analysis.

Table C: Correlation between the Presence of the Agro AS Protein
and
Tryptophan Levels in Nineteen Soy Transgenic Events bearing
pMON58120

5

Pedigree	Trp max (ppm)	Trp average (ppm)	Protein present ?
A3244 (ctr)	306	96	NO
GM_A20380:@.	6444	2246.4	YES
GM_A20532:@.	6055	2556.6	YES
GM_A22043:@.	10422	2557.2	YES
GM_A20598:@.	8861	2859.9	YES
GM_A20744:@.	7121	3373.3	YES
GM_A20381:@.	6392	3572.9	YES
GM_A20536:@.	9951	3581.5	YES
GM_A20510:@.	8916	3592.7	YES
GM_A20459:@.	8043	3900.4	YES
GM_A20337:@.	7674	4088.6	YES
GM_A20533:@.	9666	4183.2	YES
GM_A20577:@.	6276	4434.1	YES
GM_A20339:@.	9028	4687.8	YES
GM_A20386:@.	8487	5285.3	YES
GM_A20457:@.	11007	5888.9	YES
GM_A20379:@.	7672	6416.1	YES
GM_A20537:@.	9163	6695.8	YES
GM_A20534:@.	12676	7618.2	YES
GM_A20576:@.	10814	7870.1	YES

The Agro AS enzyme assay

- 10 The specific activity of anthranilate synthase was measured in eleven transformation events carrying the pMON58120 construct. Individual soybean immature seeds were analyzed using an HPLC-based end-point assay based on the method described by C. Paulsen (*J. Chromatogr.* **547**, 1991, 155-160). Briefly, desalted extracts were generated from individual seeds in grinding buffer
- 15 (100mM Tris pH7.5, 10% glycerol, 1mM EDTA, 1mM DTT) and incubated for 30 min with reaction buffer (100mM tris pH 7.5, 1 mM chorismate, 20mM glutamine, and 10mM MgCl₂). Agro AS activity was measured in the presence or absence of 25mM trp. The reaction was stopped with phosphoric acid and the

amount of anthranilate formed was quantified by HPLC using a fluorescence detector set at 340nm/excitation and 410 nm/emission.

The specific activity of AS in immature segregating transgenic seeds ranged from 1.5-fold up to 70-fold increase compared to a nontransgenic control,
5 reaching as high as 6,000 pmoles/mg/min. As shown in the last column of Table D, the anthranilate synthase activity in transgenic plants is resistant to tryptophan inhibition (see Table D).

Table D: Agro AS Enzyme Activity in Transgenic Event 20576

Event	Seed No.	Specific Activity (pmoles/mg/min)	Specific Activity (pmoles/mg/min) (+ 25 micromolar Trp)
Control	3244-1	95.4	42.4
Control	3244-2	85.5	40.6
20576	20576-1	6060.2	4407.1
20576	20576-2	3783.8	1709.4
20576	20576-3	2768.3	2431.7
20576	20576-4	4244.08	2125.2

10

**EXAMPLE 3: Soybean Transformation with a Vector
Containing a Maize Anthranilate Synthase α -Subunit gene.**

The coding sequence for a maize anthranilate synthase α -subunit was
15 isolated from pMON52214 (Figure 22) by digesting with XbaI in combination with a partial NcoI digest (see Anderson et. al. U.S. Patent 6,118,047). The resulting 1952 bp DNA fragment representing the anthranilate synthase α coding region was gel purified, and the ends were made blunt. The plasmid pMON53901 (Figure 23) was digested with BglII and EcoRI, to generate a 6.8 Kb fragment. After isolation, the ends of the 6.8 Kb fragment were made blunt and dephosphorylated. The 1952 Kb fragment containing the AS α gene was then ligated into the blunt-ended 6.8 kb pMON53901 fragment to generate
20 pMON39324, a maize 7SP-AS α -NOS expression vector (Figure 24).

This pMON39324, a maize 7SP-AS α -NOS cassette was subsequently
25 digested with BamHI resulting in a 2.84 Kb DNA fragment, containing the 7S promoter and maize AS α coding sequence. The plasmid pMON39322 (Figure 25) was digested with BamHI resulting in a 5.88 kb DNA fragment. These two

fragments were then ligated together to create pMON39325 (Figure 26), a transformation vector containing 7S promoter-maize AS α -NOS terminator cassette subcloned into pMON39322.

Using similar procedures, the coding sequence for a maize anthranilate synthase α -subunit was cloned downstream from the USP promoter to generate a pMON58130 expression vector, downstream from the Arc5 promoter to generate a pMON69662 expression vector, downstream from the Lea9 promoter to generate a pMON69650 expression vector, and downstream from the Perl promoter to generate a pMON69651 expression vector. A list with these expression vectors is presented in Table E.

Table E: C28-Maize Anthranilate Synthase Constructs

Seed Generation	Expression Cassette	Vector Name
R4	7Sa'-maize-AS α	PMON39325
R2	Napin-maize-AS α	PMON58023
R1	USP-maize-AS α	PMON58130
R1	Arc5-maize-AS α	PMON69662
R1	Lca9-maize-AS α	PMON69650
R1	Per1-maize-AS α	PMON69651

These vectors were used for plant transformation and propagation
5 experiments. Soybean plants were transformed with the maize AS-containing vectors using the microprojectile bombardment technology as described herein. Several transgenic soybean lines were established for each type of vector and propagated through the number of generations indicated in Table E.

For example, three homozygous lines were established that carried the
10 7S α -maize-AS transgene from pMON39325. These three lines were grown in a randomized block design in two different locations. Mature seed was produced and analyzed for free amino acid content. Controls were included to establish baseline trp levels, i.e. the three corresponding negative isolines and the nontransgenic controls.

15 Table F provides R4 seed tryptophan in ppm for pMON39325 transformant and control lines, showing that the average non-transgenic soybeans contain about 100-200 μ g tryptophan/g seed powder whereas the pMON39325 transformants contain substantially more Trp. *See also* Figure 27.

**Table F: Trp Levels in seeds of Soybean Plants Transformed
with the C28 *Zea mays* mutant (pMON39325)**

Positive isoline number	Average trp of Positive Isoline (ppm)	Standard deviation	Average trp of corresponding Negative isoline (ppm)	Standard deviation
39325-1	3467	377	226	55
35325-2	2623	307	164	20
35325-3	3715	152	184	64
35325-4	2833	165	202	146
35325-5	3315	161	173	34
35325-6	2394	318	144	22
nontransgenic control-7			191	24
nontransgenic control-8			118	23

Five other constructs, expressing the C28 maize anthranilate synthase under the control of five different promoters (Table E) were transformed into soy and transgenic plants were obtained. Each construct generated events high in trp. An example illustrating events generated by Perl-C28 maize anthranilate synthase is shown in Tables G and H.

10 **Table G: C28 maize AS Protein Expression Correlates
with Increased Trp Levels in Three Transgenic Events
bearing Perl-C28 maize AS (pMON69651)**

Pedigree	Trp average (ppm)	Protein present ?
Control	96	NO
22689	2375	Yes
22787	1707	Yes
22631	1116	Yes

15 Table H illustrates the enzymatic activity of C28 maize AS in R1 seeds from soybean plants transformed with the pMON69651 expression vector.

20 **Table H: Specific Activity of C28 maize AS in R1 Seeds
of pMON69651 Transformants**

Event	Seed number	Specific activity (pmoles/mg/min)	Specific activity (pmoles/mg/min) (+ 25 micromolar tryptophan)

Control		51.6	2.6
22689	22689-1	130.9	64.7
	22689-2	115.3	
	22689-3	148.5	61.1
	22689-4	149.5	
	22698-5	133.8	60.3

These results indicate that there is a substantial increase in tryptophan when soybean plant tissues are transformed with the C28 maize AS gene.

- 5 The high trp levels shown in Table G correlate with the presence of the AS protein and with increased specific activity (2.5 fold higher than in nontransgenic controls) for the transgenic enzyme (Table H). As shown in Table H - and as predicted by the biochemical properties of the C28 maize AS enzyme - the specific activity of transgenic events is tryptophan-resistant.

10

EXAMPLE 4: Rational Design of *Agrobacterium tumefaciens* Anthranilate Synthase tryptophan feedback insensitive mutants.

- This example describes vectors containing mutant *Agrobacterium tumefaciens* anthranilate synthase enzymes that have various degrees of sensitivity or insensitivity to feedback inhibition by tryptophan or tryptophan analogs.

Generation of *Agrobacterium tumefaciens* Mutant Anthranilate Synthase Genes.

- Using protein structural information from *Sulfolobus solfataricus* anthranilate synthase as a guide (Knochel et. al., *Proc. Natl. Acad. Sci. USA*, **96**, 9479-9484 (1999)) several *Agrobacterium tumefaciens* anthranilate synthase mutants were rationally designed utilizing protein informatics to confidently assign several residues involved in tryptophan binding. This was accomplished by alignment of the *Agrobacterium tumefaciens* anthranilate synthase gene with the anthranilate synthase amino acid sequence from *Sulfolobus solfataricus* (Figure 6). The putative tryptophan binding and catalysis regions of the *Agrobacterium tumefaciens* were assigned by combining the knowledge of the

structural information with the sequence homology. Residues in the binding pocket were identified as potential candidates for altering to provide resistance to feedback inhibition by tryptophan.

Based on the structural analysis of the *Sulfolobus solfataricus*

5 anthranilate synthase enzyme, it suggested that amino acids E30, S31, I32, S42, V43, N204, P205, M209, F210, G221, and A373 were involved in tryptophan binding. Based on the pairwise alignment, N204, P205, and F210 of *Sulfolobus solfataricus* were also conserved in the monomeric *Agrobacterium tumefaciens* anthranilate synthase as residues N292, P293, and F298 respectively.

10 However, due to multiple insertions and deletions, the N-terminal regions of the *Sulfolobus solfataricus* and *Agrobacterium tumefaciens* enzymes were highly divergent. For this reason, it was necessary to manually assign residues at the N-terminal region of the *Agrobacterium tumefaciens* anthranilate synthase involved in tryptophan regulation (Figure 6). Structural analysis indicated that
15 15 the motif "LLES" formed a β sheet in the tryptophan-binding pocket. This structure appeared to be highly conserved among the heterotetrameric enzymes. The known monomeric enzymes were then manually aligned to the *Sulfolobus solfataricus* sequence using the "LLES" motif as a landmark (Figure 21). Based on this protein informatics analysis, amino acid residues V48, S50, S51, and N52
20 20 in *Agrobacterium tumefaciens* AS were also likely to be involved in tryptophan binding.

With the putative tryptophan binding residues assigned in the *Agrobacterium tumefaciens* monomeric enzyme, several distinct strategies were rationalized for reducing the sensitivity of the enzyme to tryptophan inhibition.
25 25 These substitutions included for example, enlarging the tryptophan-binding pocket (F298A), narrowing the binding pocket (V48F, V48Y, S51F, S51C, N52F, F298W), increasing the polarity of the binding pocket (S50K), or distorting the shape of the binding pocket by changing the protein main chain conformation (P293A, P29G).

30

A. *tumefaciens* AS site-directed mutagenesis

Site directed mutagenesis was used to generate ten single amino acid substitutions six sites. The mutations were introduced into the *Agrobacterium*

tumefaciens AS in pMON34705 using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). The primers used for site directed mutagenesis were SEQ ID NO:9-42 (Figure 7; F = forward, R = reverse). Each primer sequence is specific for alteration of the nucleic acid at a specific location in the sequence and thus changing the encoded codon to code for a new amino acid. For example, S51C designates a change from serine to cysteine at amino acid position 51 in the *Agrobacterium tumefaciens* AS peptide sequence.

Following mutagenesis the sequence of the entire gene was reconfirmed and the variants expressed and purified from *E. coli* as described below for the wild type enzyme. The resultant plasmids comprising mutant *Agrobacterium tumefaciens* AS are suitably cloned into a plasmid for overproduction of protein using the T7 expression system as described in Example 1.

Agrobacterium tumefaciens AS protein expression and purification

Agrobacterium tumefaciens AS wild type and mutant enzymes were expressed in *E. coli* as described in Example 1. The purification of all the *Agrobacterium tumefaciens* AS enzymes, including wild type and mutants thereof, was performed at 4 °C. The cells (approximate wet weight of 1g) were suspended in 20 ml of purification buffer (50 mM potassium phosphate, pH 7.3, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 10% glycerol) and lysed by ultrasonication (Branson sonifier Cell Disruptor, W185). Supernatant was collected after centrifugation of the homogenate at 20,000 x g for 15 min. The supernatant was subjected to ammonium sulfate fractionation (30 to 65% saturation). The precipitate was collected after centrifugation at 20,000 x g for 15 min and dissolved in 3 ml of the purification buffer and then loaded as a whole on an Econo-Pac 10DG desalting column, pre-equilibrated with the same buffer. Fractions containing the enzyme were detected by the developed assay and pooled. The pooled enzyme (4.3mls) was loaded on a 10 ml DEAE Sephadex (Pharmacia Biotech) column (1.5 x 7.5 cm) equilibrated with the same buffer. The column was washed with 30 ml of the purification buffer and the enzyme was eluted with 30 ml of 50 mM NaCl in the same buffer. Fractions containing high AS activity were pooled and precipitated by 65% ammonium

sulfate saturation and isolated and desalts as above. Fractions containing the enzyme were pooled and stored at -80°C.

Anthranilate synthase enzyme assay and kinetic analysis.

- 5 The standard assay for *Agrobacterium tumefaciens* AS was performed at 25°C in an assay buffer containing 100mM potassium phosphate, pH 7.0, 10mM MgCl₂, 1mM dithiothreitol, 200μM chorismate and 10mM L-glutamine. The reaction was started by adding 30μl of enzyme to the reaction mixture and mixing. The formation of anthranilate was directly monitored by the absorbance 10 increase at 320m for 3min. Initial rate of reaction was calculated as unit absorbance increase per second based on the slope of the absorbance change over the reaction time. K_m for chorismate (K_m^{Cho}) was determined in the total volume of 1 ml assay buffer containing 100mM potassium phosphate, pH 7.0, 10mM MgCl₂, 1mM dithiothreitol with 10mM L-glutamine and varying the concentration of chorismate between 2.5-100μM chorismate. The K_m for 15 glutamine (K_m^{Gln}) was determined in the total volume of 1ml assay buffer containing 100mM potassium phosphate, pH 7.0, 10mM MgCl₂, 1mM dithiothreitol with 200μM chorismate and varying the concentration of L-glutamine between 0.1-2mM L-glutamine. IC₅₀ for tryptophan (IC₅₀^{T_{RP}}) was 20 determined with in the total volume of 1ml assay buffer containing 100mM potassium phosphate, pH 7.0, 10mM MgCl₂, 1mM dithiothreitol, 10mM L-glutamine, 200μM chorismate and varying the concentration of L-tryptophan between 0.1-10mM L-tryptophan. Kinetic parameters and IC₅₀ of AS were calculated after fitting the data to a non-linear regression program (GraFit). 25 Several mutants demonstrated reduced sensitivity to tryptophan inhibition while still maintaining enzymatic activity comparable to the wild type enzyme (Table I). These results demonstrate that the extent of sensitivity to tryptophan inhibition can be decreased, for example, by mutating amino acids in the tryptophan-binding pocket of anthranilate synthase and by optimizing of the 30 mutations demonstrating feedback insensitivity.

**Table I: Anthranilate Synthase Activity and Effect of Tryptophan
on *Agrobacterium tumefaciens* AS Mutants**

Mutation	Codon	K_m^{Cho} (μM)	K_m^{Gln} (mM)	$k_{cat}(\text{s}^{-1})$	k_{cat}/K_m^{Cho} ($\mu\text{M}^{-1}\text{s}^{-1}$)	IC_{50}^{Trp} (μM)
WT		8.0	0.11	0.43	5.37×10^{-2}	5
V48F	TTT	4.5	0.08	0.24	5.33×10^{-2}	150
V48Y	TAT	4.2	0.10	0.18	4.28×10^{-2}	650
S50K	AAG	13	0.01	0.13	1.00×10^{-2}	0.1
S51F	TTC	10	0.06	0.08	0.80×10^{-2}	>32,000
S51C	TGC	2.8	0.08	0.15	5.36×10^{-2}	1,500
N52F	TTC	5.5	0.04	0.21	3.82×10^{-2}	41
P293A	GCG	24	0.16	0.35	1.46×10^{-2}	14
P293G	GGG	33	0.07	0.48	1.45×10^{-2}	17
F298A	GCC	9.2	0.10	0.46	5.00×10^{-2}	5.5
F298W	TGG	18	0.14	0.44	2.44×10^{-2}	450

5

**EXAMPLE 5: Random mutagenesis of *Agrobacterium tumefaciens* AS
to generate tryptophan feedback insensitive mutants.**

In addition to the rational design approaches described in Example 4, other strategies to generate feedback insensitive mutants of anthranilate synthase include, but are not limited to, random mutagenesis. Random mutagenesis of the *Agrobacterium tumefaciens* AS, can be accomplished, for example, by chemical mutagenesis (isolated DNA or whole organism), error prone PCR, and DNA shuffling. This example describes the use of chemical mutagenesis followed by genetic selection. The genetic selection approach is also useful for selection of desirable mutants derived from other mutagenesis techniques.

Generation of *E. coli* expression plasmid containing *A. tumefaciens* AS

The open reading frame from the *Agrobacterium tumefaciens* AS clone pMON61600 (SEQ ID NO:1, described in Example 1) was amplified by PCR

using primers that contain an Neo 1 site on the 5' end of the forward primer and an Xba1 site on the 3' end of the reverse primer:

5'-CATCCCATGGATGGTAACGATCATT CAGGAT-3' (SEQ ID NO:55); and
5'-GATGTCTAGAGACAC TATAGAATACTCAAGC-3' (SEQ ID NO:56).

5 The resulting PCR product was ligated into pMON25997 (Figure 28), which had the bktB open reading frame (Slater et al., *J. Bact.* 180, p1979-1987 (1998)) removed by digestion with BspH1 and Xba1 resulting in plasmid pMON62000 (Figure 29). pMON62000 is the base plasmid used for mutagenesis and complementation of the tryptophan auxotroph (EMG2 Δ trpE).

10

Generation of an *E. coli* tryptophan auxotroph EMG2 Δ trpE.

E. coli strain Ec-8 (EMG2 Δ trpE) was constructed using the suicide vector pKO3 to delete 1,383 base pairs from the chromosomal *trpE* gene of *E. coli* strain EMG2(K-12 wt F+) (*E. coli* Genetic Stock Center). Two amplicons from *E. coli* genomic DNA were PCR amplified. The first amplicon was approximately 1.5kb and contained the first 30bp of the *trpE* ORF at the 3' end. This amplicon contains a BamH1 site at the 5' end and an EcoR1 site at the 3' end. The second amplicon was approximately 1kb and contained the last 150 bp of the *trpE* ORF at the 5' end. This amplicon contains an EcoR1 site at the 5' end and a SalI site at the 3' end. The two amplicons were digested with the appropriate enzymes and ligated together at the EcoR1 site to create an in-frame deletion of *trpE*. Figure 30 shows the resulting sequence of the truncated gene (SEQ ID NO:46). The *trpE* deletion amplicon was ligated into pKO3 at the BamH1 and SalI sites. Gene disruption was performed as described in A. J. Link et al. *J. Bacteriol.*, 179, 6228 (1997).

25 **Complementation of *E. coli* tryptophan auxotroph EMG2 Δ trpE with pMON62000**

E. coli strain Ec-8 (EMG2 Δ trpE) was transformed with pMON62000 and plated on M9 minimal medium to determine if the deletion was complemented by the addition of pMON62000. A plasmid control (minus the *Agrobacterium tumefaciens* AS insert) and a strain control Ec-8 were also plated onto M9 minimal medium and onto M9 minimal medium with 40 μ g/ml tryptophan.

Growth of strain Ec-8 transformed with pMON62000 was observed on M9 without tryptophan, no growth of either of the controls was observed, indicating complementation of the *trpE* deletion in strain Ec-8 by pMON62000.

5 **Hydroxylamine mutagenesis of pMON62000 and genetic selection of mutants**

To generate mutants of anthranilate synthase, pMON62000 was mutated with the chemical mutagen hydroxylamine. The following ingredients were combined in an eppendorf tube: 20 μ g pMON62000 plasmid DNA and 40 μ l 2.5

- 10 M hydroxylamine, pH 6.0. The volume was brought to a volume of 200 μ l with 0.1M NaH₂PO₄, pH 6.0 + 5mM EDTA, pH 6.0. The tube was incubated at 70°C. After 1.5 hours, 100 μ l of reaction mixture was dialyzed on a nitrocellulose filter that was floating on approximately 500ml H₂O. After 15 minutes, the DNA was concentrated using Qiagen PCR Purification Kit. After 3 hours, the remaining
- 15 100 μ l of the reaction mixture was removed and purified in the same manner.

E. coli strain Ec-8 was then transformed by electroporation with 100ng of pMON62000 that had been mutagenized for either 1.5 or 3 hours with hydroxylamine. Two transformation procedures were performed for each time point. Transformed cells were allowed to recover for 4 or 6 hours in SOC

20 medium (20g/L Bacto-Tryptone, 5g/L Bacto Yeast Extract, 10ml/L 1M NaCl, 2.5ml/L 1M KCl, 18g glucose).

- Two 245mm square bioassay plates were prepared containing M9 minimal medium, plus 2% agar, and 50 μ g/ml 5-methyl-DL-tryptophan (5-MT). An aliquot of 900 μ l of the 1.5 hour mutagenized transformation mixture was
- 25 plated onto one 50 μ g/ml 5-MT plate. The remaining 100 μ l was plated onto the M9 control plate. The same procedure was performed for the transformation mixture containing the 3.0 hour mutagenized plasmid.

- The plates were then incubated at 37°C for approx. 2.5 days. Resistant colonies were isolated from the 5-MT plates and were streaked onto LB-
- 30 kanamycin (50 μ g/ml) plates to confirm the presence of the plasmid. All of the selected colonies grew on these plates. Individual colonies from each of the resistant clones were prepped in duplicate to isolate the plasmid. Restriction

digests and PCR were performed and confirmed that all the clones contained the desired *Agrobacterium tumefaciens* AS insert.

- The rescued plasmids were then transformed back into strain Ec-8. One colony from each transformation was purified by streaking onto new LB-5 Kanamycin plates. To confirm resistance to 5-MT, individual purified colonies were streaked onto plates containing M9 plus 50 µg/ml 5-MT and 2% agar, and then grown at 37°C for 3 days. Resistance was confirmed for most of the clones. To determine if resistant mutants would remain resistant at an even higher concentration of 5-MT, they were plated onto M9 plus 300 µg/ml 5-MT and 2% Agar. Most clones demonstrated resistance at this high concentration also.

The plasmids from all of the resistant clones were isolated and sequenced on both strands. Some of the mutations from this experiment are diagrammed in Table J.

15

Table J: *A. tumefaciens trpEG Sequence Variations in 5-MT Resistant Clones.*

Database Clone #	Original Clone #	Determined Sequence Variations	K _m ^{cho} (µM)	IC ₅₀ ^{trp} (µM)
Wt			8.0	5.0
Ec-12	1	G4A Val2Ile		
Ec-18	8	C35T Thr12Ile	15	2.5
Ec-19	9	C2068T Pro690Ser	5.0	3.4
Ec-20	11	G1066A Glu356Lys & C1779T Ile593Ile		

- As indicated by the data in Table J, several mutants had little effect on the K_m and IC₅₀ of the mutant enzyme, indicating that these mutations are likely not the source of resistance to tryptophan feedback inhibition. For example, the mutation of C to T at nucleotide 35, which changes a threonine residue to isoleucine at amino acid position 12 (Thr12Ile), gives rise to a minor change in K_m^{cho} and IC₅₀^{trp} values. Similarly, a change of C to T at nucleotide position 2068, which changes a proline to a serine also gives rise to a minor change in K_m^{cho} and IC₅₀^{trp} values. These mutations may therefore, may be "silent"

mutations that give rise to variant gene products having enzymatic properties like those of wild type.

EXAMPLE 6: High Tryptophan Transgenic Soybean Plants.

5 This example sets forth preparation of transgenic soybean plants having elevated tryptophan levels resulting from transformation with tryptophan feedback insensitive mutants of anthranilate synthase from *Agrobacterium tumefaciens*.

10 Vector Construction

Plasmid pMON34711, which harbors the anthranilate synthase clone from *Agrobacterium tumefaciens* containing the F298W mutation described in Example 4, was digested with restriction enzyme NotI. The ends of the resulting fragment were blunted and then digested with NcoI. The plasmid pMON13773 (Figure 8) was then digested with restriction enzyme EcoRI, the ends blunted and then digested with NcoI. The resulting fragments were ligated resulting in plasmid pMON58044, which contained the AS gene under the control of the 7S promoter and NOS3' terminator (Figure 9).

20 Plasmid pMON58044 was then cut with restriction enzymes BglII and NcoI and ligated with a fragment that was generated by digesting pMON53084 (Figure 10) with BglII and NcoI. The resulting fragment was named pMON58045 (Figure 11) and contained the sequence for the *Arabidopsis* SSU1A transit peptide.

Finally, plasmid pMON58046 (Figure 12) was constructed by ligating the 25 fragments generated by digesting pMON58045 (Figure 11) and pMON38207 (Figure 13) with restriction enzyme NotI. This resulted in the pMON58046 vector (Figure 12) that was used for soybean transformation.

Soybean Transformation By Microprojectile Bombardment

30 For the particle bombardment transformation method, commercially available soybean seeds (i.e., Asgrow A3244, A4922) were germinated overnight for approximately 18-24 hours and the meristem explants were excised. The primary leaves were removed to expose the meristems and the explants were

placed in targeting media with the meristems positioned perpendicular to the direction of the particle delivery.

The pMON58046 transformation vector described above was precipitated onto microscopic gold particles with CaCl₂ and spermidine and subsequently 5 resuspended in ethanol. The suspension was coated onto a Mylar sheet that was then placed onto the electric discharge device. The particles were accelerated into the plant tissue by electric discharge at approximately 60% capacitance.

Following bombardment, the explants were placed in selection media (WPM + 0.075 mM glyphosate) (WPM = Woody Plant Medium (McCown & 10 Lloyd, Proc. International Plant Propagation Soc., 30:421, 1981) minus BAP)) for 5-7 weeks to allow for selection and growth of transgenic shoots. Phenotype positive shoots were harvested approximately 5-7 weeks post-bombardment and placed into selective rooting media (BRM + 0.025mM glyphosate) (see below for BRM recipe) for 2-3 weeks. Shoots producing roots were transferred to the 15 greenhouse and potted in soil. Shoots that remained healthy on selection, but did not produce roots were transferred to non-selective rooting media (BRM without glyphosate) for an additional two weeks. The roots from any shoots that produced roots off the selection were tested for expression of the plant selectable marker before transferring to the greenhouse and potting in soil. Plants were 20 maintained under standard greenhouse conditions until R1 seed harvest.

The recipe used for Bean Rooting Medium (BRM) is provided below.

	<u>Compound</u>	<u>Quantity for 4L</u>
	MS Salts***	8.6g
25	Myo-inositol(cell culture grade)	0.40g
	SBRM Vitamin Stock**	8.0ml
	L-Cysteine (10mg/ml)	40.0ml
	Sucrose (ultra pure)	120g
	Adjust pH to 5.8	
30	Washed Agar	32g
	Additions after autoclaving:	
	SBRM/TSG Hormone Stock*	20.0ml

- *SBRM/TSG Hormone Stock (per 1L of BRM): 3.0ml IAA (0.033mg/ml),
2.0ml sterile distilled water. Store stock in dark at 4 °C.
- **SBRM Vitamin Stock (per 1L of stock): Glycine (1.0g), Nicotinic Acid (0.25g), Pyridoxine HCl (0.25g), Thiamine HCl (0.25g).
- 5 ***3X MInor MS Salts (per 1L stock): H₂BO₃ (1.86g), MnSO₄ (5.07g), ZnSO₄·H₂O (2.58g), KI (0.249g), 7.5 ul NaMoO₄·2H₂O (1.0mg/ml), 7.5 ul CoSO₄·5H₂O (1.0mg/ml), 7.5 ul CoCl₂·6H₂O (1.0mg/ml). One ingredient at a time was added and dissolved, the volume was brought to one liter with sterile distilled water, and the solution was stored in a foil-covered bottle in the refrigerator for no longer than one month.
- 10

Soybean Transformation Using *Agrobacterium tumefaciens*

- For the *Agrobacterium* transformation method, commercially available soybean seeds (Asgrow A3244, A4922) were germinated overnight
- 15 (approximately 10-12 hours) and the meristem explants were excised. The primary leaves may or may not have been removed to expose the meristems and the explants were placed in a wounding vessel.

Agrobacterium strain ABI containing the plasmid of interest was grown to log phase. Cells were harvested by centrifugation and resuspended in

20 inoculation media containing inducers. Soybean explants and the induced *Agrobacterium* culture were mixed no later than 14 hours from the time of initiation of seed germination and wounded using sonication.

Following wounding, explants were incubated in *Agrobacterium* for a period of approximately one hour. Following this inoculation step, the

25 *Agrobacterium* was removed by pipetting and the explants were placed in co-culture for 2-4 days. At this point, they were transferred to selection media (WPM + 0.075 mM glyphosate + antibiotics to control *Agrobacterium* overgrowth) for 5-7 weeks to allow selection and growth of transgenic shoots.

Phenotype positive shoots were harvested approximately 5-7 weeks post-30 bombardment and placed into selective rooting media (BRM + 0.025 mM glyphosate) for 2-3 weeks. Shoots producing roots were transferred to the greenhouse and potted in soil. Shoots that remained healthy on selection, but did not produce roots were transferred to non-selective rooting media (BRM without

glyphosate) for an additional two weeks. The roots from any shoots that produced roots off the selection were tested for expression of the plant selectable marker glyphosate resistance before transferring to the greenhouse and potting in soil. Plants were maintained under standard greenhouse conditions until R1 seed 5 harvest.

Analysis of Amino Acid Content of R1 Seed

Mature R1 seed is produced and analyzed for free amino acid content using fluorescence detection as described in Agilent Technologies Technical 10 Bulletin REV14. Five seeds are chosen for single seed analysis from each event. Soy seeds expressing the AgroAS F298W or the AgroAS S51F mutant proteins generate very high amounts of tryptophan. Results are shown in Tables K and L.

Table K: Protein expression in Seeds Transformed with pMON58046

Pedigree	Trp average (ppm)	Protein present ?
Control	96	no
22817	9922	yes
22891	12955	yes
23026	7968	yes

5

Table L: AS Protein expression Correlated with pMON58123 Transformation

Pedigree	Trp average (ppm)	Protein present ?
Control	96	no
23562	88	no
23590	8795	yes
23911	388	no

10

AS Enzyme Activity in R1 Seed Transformed with Agro AS

Mature R1 seed is produced and analyzed for anthranilate synthase activity. Anthranilate synthase enzymatic activity was determined in R1 soy seeds carrying the AgroAS F298W (SEQ ID NO:65 or 91) or the Agro AS S51F (SEQ ID NO:60 or 86) mutant alleles. Very high levels of tryptophan-resistant anthranilate synthase activity was observed, consistent with the high amounts of tryptophan generated by these seeds. Results are shown in Tables M and N.

20

Table M: Specific activity of AS in R1 Seeds Transformed with pMON58046

Event	Seed number	Specific activity (pmoles/mg/min)	Specific activity (pmoles/mg/min) (+ 25 micromolar Trp)
Control		77.6	
23076	23076-1	100.5	1.04
	23076-2	4512.8	
	23076-3	9737.4	9290.4
	23076-4	136.12	
	23076-5	8992.5	9749.9

**Table N: Specific activity of AS in R1 Seeds
Transformed with pMON58123**

Event	Seed number	Specific activity (pmoles/mg/min)	Specific activity (pmoles/mg/min) (+ 25 micromolar Trp)
Control		83.7	32.7
23590	23590-1	891	692.3
	23590-2	466.2	186.5
	23590-3	71.7	38.3
	23590-4	320.5	316.2

5

EXAMPLE 7: Preparation of Transformation Vector Comprising *Ruta graveolens* Anthranilate Synthase α -Subunit

- The anthranilate synthase α gene from *Ruta graveolens* (Genbank Accession No. GI 960291) provides another anthranilate synthase domain useful in the present invention (Bohlmann, J et al., *Plant Phys* 111 507-514 (1996)). One isoenzyme of anthranilate synthase present in the genome of *Ruta graveolens* demonstrates less susceptibility to feedback inhibition by L-tryptophan. This allele may also be useful in the present invention to elevate the levels of free L-tryptophan in transgenic plants. The vector pMON58030 (Figure 14) contains the *Ruta graveolens* anthranilate synthase α -subunit that is less sensitive to tryptophan inhibition. The *Ruta graveolens* anthranilate synthase α gene was PCR amplified from pMON58030 to provide a BamHI site at the 5' end and a BgIII site at the 3' end of the *Ruta graveolens* anthranilate synthase α gene fragment by utilizing PCR primers that contained these two restriction enzyme sites:
- 5'-CAAAAGCTGGATCCCCACC-3' (SEQ ID NO:53) and
5'-CCTATCCGAGATCTCTCACTCC-3' (SEQ ID NO:54).
- The PCR fragment was purified, digested with the respective restriction enzymes, to form pMON58041, which contains the transcriptional fusion of the *Ruta graveolens* AS α to the napin promoter. The *Agrobacterium* mediated plant transformation plasmid, pMON58043, was created comprising the napin promoter, *Ruta graveolens* AS, NOS terminator, glyphosate resistance (CP4)

selectable marker and borders suitable for proper chromosomal integration of the cassette as described. The resulting plant transformation vector was used to transform plants using standard plant transformation techniques as described in Examples 2, 3 and 6.

5

EXAMPLE 8: Transforming multi-polypeptide anthranilate synthases into monomeric single polypeptide anthranilate synthases

Generation of a monomeric anthranilate synthase by fusion of selected multi-subunit enzymes is desirable, for example, to maximize the catalytic efficiency, to stabilize the enzyme, to achieve coordinated expression, for example, of subunits comprising activities of TrpE and TrpG and for effective communication between the two subunits. In some instances, it may be useful to employ TrpE or α -subunits from either plant or microbial source that are deregulated with respect to feedback inhibition by standard mutagenesis techniques or by rational design as described in the foregoing Examples, e.g. in Example 4. In other instances, wild type TrpE or α -subunits from either plant or microbial source are employed.

The C-terminus of the selected TrpE or α -subunit is linked to the N-terminus of the TrpG subunit or β -subunit, preferably with a peptide linker. A linker can be rationally designed to provide suitable spacing and flexibility for both subunits to properly align. Alternatively a linker can be identified by sequence alignment of monomeric and heterotetrameric anthranilate synthases. Examples of sequence alignments of monomeric and heterotetrameric anthranilate synthase forms are shown in Figures 21 and 35. It is also envisioned that it may be necessary to generate monomeric anthranilate synthases comprising heterologous subunit in order to maximize the benefits. For example, an α -subunit may be obtained from a bacterial source, for example, *E. coli* and fused to a β -subunit from a plant source, for example, *Arabidopsis*.

The novel protein produced can be introduced into plants, for example, as described in Examples 2, 3 or 6. The invention is not limited to the exact details shown and described, for it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention defined by the claims.

**EXAMPLE 9: Identification of anthranilate synthases
from genomic sequence databases.**

5 Monomeric anthranilate synthases as well as α and β domains useful in the invention can be identified by bioinformatics analysis by searching for example, genbank and/or swissprot databases using BLAST (www.ncbi.nlm.nih.gov/blast/). Useful query sequences to identify monomeric anthranilate synthase include, for example, domains of anthranilate synthase 10 such as the α -domain (GI 1004323) or β -domain (GI 1004324) from *Sulfolobus solfataricus*, or monomeric anthranilate synthase such as *Agrobacterium tumefaciens* AS (GI 15889565). Putative monomeric anthranilate synthase will have between 50% and 100% homology with the query sequence and should minimally contain 700 amino acids. If the AS- α -domain is used to query the 15 genomic database, in addition to identifying putative anthranilate synthase genes it is also likely to identify genes involved in PABA synthesis for example 4-amino-4-deoxychorismate (ADC) synthase. The monomeric ADC synthase genes can be easily identified away from putative monomeric AS genes based on the observation that the amidotransferase domain (β -domain) of ADC synthase 20 resides at the N-terminus of the protein whereas the amidotransferase domain (β -domain) of AS resides at the C-terminus. Monomeric anthranilate synthases useful in the present invention identified by bioinformatics analysis include, but are not limited to, for example, *Rhizobium meliloti* (GI 95177), *Mesorhizobium loti* (GI 13472468), *Brucella melitensis* (GI 17982357), *Nostoc sp. PCC7120* (GI 25 17227910, GI 17230725), *Azospirillum brasilense* (GI 1174156), *Rhodopseudomonas palustris*, *Anabaena M22983* (GI 152445). Figure 21 is an example of a sequence alignment of two monomeric anthranilate synthases (*Agrobacterium tumefaciens* and *Rhizobium meliloti*) with two heterotetrameric anthranilate synthases (*Sulfolobus solfataricus* and *Arabidopsis thaliana*) useful 30 in the present invention. Figure 35 is an example of a sequence alignment of several monomeric anthranilate synthases with the *Rhodopseudomonas palustris* heterotetrameric anthranilate synthase.

EXAMPLE 10: Optimized Codon Usage

This example sets forth a method of improving the expression of an anthranilate synthase gene in the seed of a plant by optimization of the codon usage.

5 The nucleotide sequence of the anthranilate synthase (AS) gene from wild type *Agrobacterium tumefaciens* (SEQ ID NO:1) was inspected for the presence of underexpressed codons. To identify underexpressed codons sequences of highly expressed seed proteins from corn and soybeans were examined for relative codon frequency. The relative codon usage frequencies are
10 shown in Table O represented in an expected value format. Expected value format can be exemplified as follows: Assume there are four codons that encode a given amino acid, and assume that they are used equally well, then each codon would be expected to account for 25% (0.25) of the frequency for that amino acid. However, due to redundancy, 0.25 was normalized to 1.0 to give a relative
15 score for each codon as compared to other codons that encode that amino acid. For this analysis, if a codon was more prevalent than the other choices for a given amino acid, it received a number that was greater than 1.0. Correspondingly, if a codon was less prevalent, it received a number less than 1.0. For this study, a particular codon was considered underrepresented if it's relative codon usage
20 frequency was lower than 0.5.

Using the results from Table O, a close examination of the wild type *Agrobacterium* AS sequence revealed that 125 codons were considered underrepresented (below the threshold of 0.5) in corn and soybeans (Table P). These underrepresented codons were replaced by more prevalent codons as
25 defined above. The modified nucleotide sequence is shown in Figure 36. Using bioinformatics tools, the resulting sequence was assembled and analyzed for integrity by translation and alignment of the nucleotide and protein sequences with the corresponding wild type AS sequences. While, the protein sequence was unchanged the nucleotide sequence of the optimized sequence had 94%
30 identity with the wild type *Agrobacterium* AS sequence (Figure 37). The optimized nucleotide sequence was analyzed for the absence of cryptic polyadenylation signals (AATAAA, AATAAT) and cryptic introns using Lasergene EditSeq (DNASTAR, Inc., Madison, WI) and Grail2 (Oak Ridge

National Laboratory, Oak Ridge, TN), respectively. No cryptic signals were found.

The modified nucleotide sequence is synthesized using techniques well known in the art or by commercial providers such as Egea Biosciencesces, Inc.

- 5 (San Diego, CA). The resulting nucleotide is cloned into an appropriate expression vector and tested for efficacy in corn, soybeans and *Arabidopsis* using procedures detailed in earlier examples of this specification.

**Table O: Relative codon usage frequencies
in maize and soybean seed-expressed genes¹.**

Codon	AA	Maize Seed	Soy Seed	Codon	AA	Maize Seed	Soy Seed
TTT	F	0.4211	0.7348	ATC	I	1.7143	1.0563
TTC	F	1.5789	1.2652	ATA	I	0.3673	0.6654
TTA	L	0.4557	0.3875	ATG	M	1.0000	1.0000
TTG	L	0.9494	1.2060	ACT	T	0.6153	1.0008
TCT	S	0.9624	1.4851	ACC	T	1.2213	2.1020
TCC	S	1.3707	1.1249	ACA	T	0.8372	0.7146
TCA	S	0.9107	1.0044	ACG	T	1.3262	0.1826
TCG	S	0.7851	0.3266	AAT	N	0.2885	0.5409
TAT	.Y	0.2455	0.6861	AAC	N	1.7115	1.4591
TAC	Y	1.7545	1.3139	AAA	K	0.5333	0.9030
TGT	C	0.2778	0.7572	AAG	K	1.4667	1.0970
TGC	C	1.7222	1.2428	AGT	S	0.2679	0.9714
TGG	W	1.0000	1.0000	AGC	S	1.7032	1.0876
CTT	L	0.7975	1.6298	AGA	R	0.3913	1.9459
CTC	L	1.0610	1.6301	AGG	R	2.9185	1.3087
CTA	L	0.8544	0.5905	GTT	V	0.5714	1.2381
CTG	L	1.8820	0.5562	GTC	V	1.0119	0.6864
CCT	P	0.6500	1.5822	GTA	V	0.3810	0.3472
CCC	P	0.8520	0.7694	GTG	V	2.0357	1.7284
CCA	P	1.2240	1.5838	GCT	A	0.9876	1.3583
CCG	P	1.2740	0.0645	GCC	A	1.1618	1.1283
CAT	H	0.8438	0.6066	GCA	A	0.8011	1.2898
CAC	H	1.1563	1.3934	GCG	A	1.0495	0.2235
CAA	Q	0.8639	1.2162	GAT	D	0.8500	0.9523
CAG	Q	1.1361	0.7838	GAC	D	1.1500	1.0477
CGT	R	0.2582	0.5903	GAA	E	0.6818	1.0463
CGC	R	1.0082	1.1159	GAG	E	1.3182	0.9537
CGA	R	0.1957	0.6700	GGT	G	1.1268	1.1431
CGG	R	1.2283	0.3692	GGC	G	1.8758	0.8577
ATT	I	0.9184	1.2783	GGA	G	0.3085	1.2759
ATC	I	1.7143	1.0563	GGG	G	0.6889	0.9233

- 5 ¹ The relative codon frequencies are represented in the expected value format. This means
that if there are four codons that encode a given amino acid, and they are used equally well,
each codon is expected to account for 25% (0.25). Due to the redundancy, 0.25 was
normalized to 1 to give a relative score for each codon as compared to all codons that encode
that amino acid. In real life if a codon is more prevalent than the other choices for a given
10 amino acid, it would get a number >1. And if it is less preferred than the other codons for the
amino acid, it would get a number <1.

Table P: Underrepresented Agro AS codons and modifications for improved seed expression^a.

Codon	Codon	Amino	Modified	Underrep	Codon	Codon	Amino	Modified	Underrep	Codon	Codon	Amino	Modified	Underrep
(wt)	Acid	Codon	In Crop ²	(wt)	Acid	Codon	In Crop	(wt)	Acid	Codon	Codon	Amino	Modified	Underrep
2	GTA	V	GTC	com, soy	177	TCG	S	TCC	soy	481	GCG	A	GCC	soy
3	ACG	T	GGC	soy	179	GCG	A	GCC	soy	485	AAT	N	AAC	com, soy
9	GGA	G	GGT	com	180	CGT	R	GCG	com	489	CCG	P	CCA	soy
10	GCG	A	GCC	soy	181	CCG	P	CCA	soy	504	ATA	-	ATC	com
15	ACG	T	ACC	soy	185	CGT	R	GCG	com	508	CGT	R	CGC	com
16	AAA	K	AAG	com	190	TTT	F	TTC	com	520	CGT	R	CGC	com
21	GTC	V	GTC	soy	201	TAT	Y	TAC	com	543	ACG	T	ACC	soy
23	CGA	R	CGC	com	209	CGT	R	GCG	com	545	GCG	A	GCC	soy
26	CGG	R	CGC	soy	218	ACG	T	ACC	soy	546	AAT	N	AAC	com, soy
30	TAT	Y	TAC	com	219	ACG	T	ACC	soy	547	TAT	Y	TAC	com
36	AAT	N	AAC	com, soy	238	CCG	P	CCA	soy	551	ACG	T	ACC	soy
46	GGC	G	GGT	soy	244	CGT	R	GCG	com	553	GCG	A	GCC	soy
47	GCG	A	GCC	soy	248	TAT	Y	TAC	com	554	ACG	T	ACC	soy
48	GTT	V	GTG	com	276	CGT	R	GCG	com	556	TCG	S	TCC	soy
49	TTT	F	TTC	com	280	AAT	N	AAC	com, soy	559	AGA	R	AGG	com
50	TCG	S	TCC	soy	281	CCG	P	CCA	soy	561	CCG	P	CCA	soy
53	TAT	Y	TAC	com	282	TCG	S	TCC	soy	572	CCG	P	CCA	soy
55	TAT	Y	TAC	com	283	GCG	A	GCC	soy	578	TCG	S	TCC	soy
56	CCG	P	CCA	soy	290	GCG	A	GCC	soy	580	GGG	G	GGT	com
58	CGT	R	CGC	com	293	CCG	P	CCA	soy	584	CCG	P	CCA	soy

	Codon	Codon	Amino	Modified	Underrep	Codon	Codon	Amino	Modified	Underrep	Codon	Codon	Amino	Modified	Underrep
(wt)	Acid	Codon	in Crop ²	(wt)	Acid	Codon	Codon	in Crop	(wt)	Acid	Codon	Codon	in Crop	(wt)	Acid
64	ACG	T	ACC	294	TGT	S	TCC	soy	585	ACG	T	ACC	Soy		
69	CCG	P	CCA	296	TAT	Y	TAC	corn	592	CCG	T	CCA	Soy		
70	CCG	P	CCA	301	AAT	N	AAC	corn, soy	602	CCG	P	CCA	Soy		
75	TGT	C	TGC	307	TAT	Y	TAC	corn	617	TAT	Y	TAC	Corn		
76	TTT	F	TTG	312	TCG	S	TCC	soy	633	TCG	S	TCC	Soy		
85	TAT	Y	TAC	313	CGG	P	CCA	soy	652	ACG	T	ACC	Soy		
86	AAT	N	AAC	322	CGT	R	CGC	corn	655	CGT	R	CGC	Corn		
97	ACG	T	ACC	328	CCG	P	CCA	soy	658	TCG	S	TCC	Soy		
102	GCG	A	GCC	329	ATA	I	ATC	corn	667	CCG	P	CCA	Soy		
112	TCG	S	TCC	339	CCG	P	CCA	soy	668	CGT	R	CGC	Corn		
115	CGG	R	CGC	352	TGG	S	TCC	soy	680	ACG	T	ACC	Soy		
123	CCG	P	CCA	363	TGG	S	TCC	soy	690	CCG	P	CCA	Soy		
125	CGT	R	CGC	376	CCG	P	CCA	soy	698	CCG	P	CCA	Soy		
133	TCG	S	TCC	378	TGG	S	TCC	soy	700	TCG	S	TCC	Soy		
136	CGG	P	CCA	390	TAT	Y	TAC	corn	703	ACG	T	ACC	Soy		
137	ACG	T	ACC	411	TTT	F	TTC	corn	705	GGA	G	GGT	Corn		
143	AGA	R	AGG	442	CCG	P	CCA	soy	708	GCG	A	GCC	Soy		
150	TAT	Y	TAC	446	TAT	Y	TAC	corn	711	GCG	R	CGC	Soy		
151	TCG	S	TCC	449	GCG	A	GCC	soy	715	AAT	N	AAC	corn, soy		
153	GCG	A	GCC	460	AAT	N	AAC	corn, soy	724	GCG	A	GCC	Soy		
155	TCG	S	TCC	464	ACG	T	ACC	soy	729	GCG	A	GCC	Soy		
173	GCG	A	GCC	469	CGG	R	CGC								

²The columns titled "Underrep in Crop" indicate in which crop (maize or soybean) a particular codon is underrepresented.

All publications and patents are incorporated by reference herein, as though individually incorporated by reference. The invention is not limited to the exact details shown and described, for it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention defined by the claims.

WHAT IS CLAIMED:

1. An isolated DNA encoding a monomeric anthranilate synthase, wherein the monomeric anthranilate synthase comprises a single polypeptide comprising an anthranilate synthase α -domain and an anthranilate synthase β -domain, and wherein
5 the monomeric anthranilate synthase is expressed in a plant.
2. The isolated DNA of claim 1, wherein expression of the monomeric anthranilate synthase elevates the level of L-tryptophan in the plant relative to an untransformed plant having the same genetic background.
10
3. The isolated DNA of claim 1, wherein the monomeric anthranilate synthase is an *Agrobacterium tumefaciens*, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc sp.* PCC7120, *Azospirillum brasilense* or *Anabaena M22983* anthranilate synthase.
15
4. The isolated DNA of claim 1, wherein the monomeric anthranilate synthase comprises any one of SEQ ID NO:4, 7, 43, 58, 59, 60, 61, 62, 63, 64, 65, 69, 70, 77, 78, 79, 80, 81 or 82.
- 20 5. The isolated DNA of claim 1, wherein the isolated DNA comprises any one of SEQ ID NO:1, 75, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92 or 93.
6. The isolated DNA of claim 1, wherein the isolated DNA encodes a chimeric monomeric anthranilate synthase comprising a fusion of an anthranilate synthase α domain from one species and an anthranilate synthase β domain from a second species.
25

7. The isolated DNA of claim 1, wherein DNA encoding the α domain or the β domain is obtained from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, cotton, rice, wheat, tobacco or *Zea mays*.
5
8. The isolated DNA of claim 1, wherein the α domain or the β domain is at least a portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45, 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81, 82, 99, 100, 101, 102 or 103.
10
9. The isolated DNA of claim 1, wherein the anthranilate synthase comprises a mutation that increases anthranilate synthase activity or reduces the sensitivity of the anthranilate synthase to inhibition by tryptophan or an analog thereof.
15
10. The isolated DNA of claim 9, wherein the mutation is in a tryptophan-binding pocket.
20
11. The isolated DNA of claim 9, wherein the mutation is within amino acid positions 25-60 or 200-225 or 290-300 or 370-375 when the anthranilate synthase amino acid sequence is aligned with a monomeric *Agrobacterium tumefaciens* anthranilate synthase having SEQ ID NO:4.
25
12. The isolated DNA of claim 9, wherein the mutation is:
 - (a) at about position 48, replace Val with Phe;
 - (b) at about position 48, replace Val with Tyr;

- (c) at about position 51, replace Ser with Phe;
(d) at about position 51, replace Ser with Cys;
(e) at about position 52, replace Asn with Phe;
(f) at about position 293, replace Pro with Ala;
5 (g) at about position 293, replace Pro with Gly; or
(h) at about position 298, replace Phe with Trp; and
wherein the position of the mutation is determined by alignment of the amino acid
sequence of the anthranilate synthase with an *Agrobacterium tumefaciens*
anthranilate synthase amino acid sequence.
- 10 13. The isolated DNA of claim 9, wherein the anthranilate synthase comprises any one
of SEQ ID NO:58-65, 69 or 70.
14. The isolated DNA of claim 12, wherein the *Agrobacterium tumefaciens* anthranilate
15 synthase amino acid sequence is SEQ ID NO:4.
16. The isolated DNA of claim 1, wherein the isolated DNA further encodes a plastid
transit peptide.
- 20 17. The isolated DNA of claim 1, wherein the isolated DNA further encodes a
selectable marker gene or a reporter gene.
- 25 18. The isolated DNA of claim 17, wherein the selectable marker gene, when expressed
in a plant, imparts herbicide resistance to cells of said plant.

19. The isolated DNA of claim 18, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.
20. The isolated DNA of claim 1, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.
21. The isolated DNA of claim 1, wherein the plant is a dicot.
- 10 22. The isolated DNA of claim 21, wherein the plant is soybean or canola.
23. The isolated DNA of claim 1, wherein the plant is a monocot.
24. The isolated DNA of claim 23, wherein the plant is maize, rice, wheat, barley or sorghum.
- 15 25. The isolated DNA of claim 1, wherein the isolated DNA encoding the anthranilate synthase comprises a promoter operably linked thereto.
- 20 26. A vector comprising the isolated DNA of any one of claims 1- 25.
27. A seed comprising the isolated DNA of any one of claims 1- 25.
- 25 28. A transgenic plant comprising an isolated DNA encoding a monomeric anthranilate synthase operably linked to a promoter, wherein the monomeric anthranilate synthase comprises an anthranilate synthase α domain and an anthranilate synthase β domain, and wherein the monomeric anthranilate synthase is expressed in the plant.

29. The transgenic plant of claim 28, wherein expression of the monomeric anthranilate synthase elevates the level of L-tryptophan in the plant relative to an untransformed plant having the same genetic background.

5

30. The transgenic plant of claim 28, wherein the monomeric anthranilate synthase is an *Agrobacterium tumefaciens*, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc* sp. PCC7120, *Azospirillum brasiliense* or *Anabaena* M22983 anthranilate synthase.

10

31. The transgenic plant of claim 30, wherein the monomeric anthranilate synthase is a *Rhizobium meliloti* (Genbank Accession No. GI 95177), *Mesorhizobium loti* (Genbank Accession No. GI 13472468), *Brucella melitensis* (Genbank Accession No. GI 17982357), *Nostoc* sp. PCC7120 (Genbank Accession No. GI 17227910, GI 17230725), *Azospirillum brasiliense* (Genbank Accession No. GI 1174156) or *Anabaena* M22983 (Genbank Accession No. GI 152445) anthranilate synthase.

15

32. The transgenic plant of claim 28 wherein the monomeric anthranilate synthase is a chimeric monomeric anthranilate synthase comprising a fusion of an anthranilate synthase α domain from one species linked to an anthranilate synthase β domain from a second species.

20

33. The transgenic plant of claim 28, wherein DNA encoding the α domain or the β domain is obtained from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*,

25

Serratia marcescens, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco or *Zea mays*.

34. The transgenic plant of claim 28, wherein the α domain or the β domain is at least a portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45, 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81, 82, 99, 100, 101, 102 or 103.

35. The transgenic plant of claim 28, wherein the anthranilate synthase comprises a mutation that increases anthranilate synthase activity or reduces the sensitivity of the anthranilate synthase to inhibition by tryptophan or an analog thereof.

36. The transgenic plant of claim 35, wherein the mutation is in a tryptophan-binding pocket.

37. The transgenic plant of claim 35, wherein the mutation is within amino acid positions 25-60 or 200-225 or 290-300 or 370-375 when the anthranilate synthase amino acid sequence is aligned with a monomeric *Agrobacterium tumefaciens* anthranilate synthase having SEQ ID NO:4.

38. The transgenic plant of claim 35, wherein the mutation is:

 - at about position 48, replace Val with Phe;
 - at about position 48, replace Val with Tyr;
 - at about position 51, replace Ser with Phe;
 - at about position 51, replace Ser with Cys;
 - at about position 52, replace Asn with Phe;
 - at about position 293, replace Pro with Ala;
 - at about position 293, replace Pro with Gly; or

(h) at about position 298, replace Phe with Trp; and
wherein the position of the mutation is determined by alignment of the amino acid sequence of the anthranilate synthase with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence.

5

39. The transgenic plant of claim 35, wherein the anthranilate synthase comprises any one of SEQ ID NO:58-65, 69 or 70.
40. The transgenic plant of claim 38, wherein the *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence is SEQ ID NO:4.
10
41. The transgenic plant of claim 28, wherein the isolated DNA further comprises a plastid transit peptide.
42. The transgenic plant of claim 41, wherein the plastid transit peptide comprises SEQ ID NO:72 or 74.
15
43. The transgenic plant of claim 28, wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.
44. The transgenic plant of claim 43, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of said plant.
20
45. The transgenic plant of claim 44, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.
25

46. The transgenic plant of claim 28, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.
- 5 47. The transgenic plant of claim 28, wherein the plant is a dicot.
48. The transgenic plant of claim 47, wherein the plant is soybean or canola.
- 10 49. The transgenic plant of claim 28, wherein the plant is a monocot.
50. The transgenic plant of claim 49, wherein the plant is maize, rice, wheat, barley or sorghum.
- 15 51. A seed of the transgenic plant of claim 28.
52. A transgenic plant comprising, operably linked to a promoter, an isolated DNA encoding an *Agrobacterium tumefaciens* anthranilate synthase, or a domain thereof.
- 20 53. The transgenic plant of claim 52, wherein the *Agrobacterium tumefaciens* anthranilate synthase, or domain thereof, is expressed so as to elevate the level of L-tryptophan in said plant.
- 25 54. The transgenic plant of claim 52, wherein the *Agrobacterium tumefaciens* anthranilate synthase comprises SEQ ID NO:4, 58, 59, 60, 61, 62, 63, 64, 65, 69 or 70.
55. The transgenic plant of claim 52, wherein the isolated DNA comprises SEQ ID NO:1, 75, 84, 85, 86, 87, 88, 89, 90, 91, 92 or 93.

56. A transgenic plant comprising, operably linked to a promoter, an isolated DNA encoding a chimeric monomeric anthranilate synthase, wherein the anthranilate synthase is a fusion of an anthranilate synthase α domain from one species and an anthranilate synthase β domain from a second species.
57. The transgenic plant of claim 56, wherein the chimeric monomeric anthranilate synthase is expressed so as to elevate the level of L-tryptophan in said plant.
- 10 58. The transgenic plant of claim 56, wherein DNA encoding the α domain or the β domain is obtained from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco or *Zea mays*.
- 15 59. The transgenic plant of claim 56, wherein the α domain or the β domain is at least a portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45, 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81 , 82, 99, 100, 101, 102 or 103.
- 20 60. The transgenic plant of claim 52 or 56, wherein the anthranilate synthase comprises a mutation that increases anthranilate synthase activity or reduces the sensitivity of the anthranilate synthase to inhibition by tryptophan or an analog thereof.
- 25 61. The transgenic plant of claim 60, wherein the mutation is within amino acid positions 25-60 or 200-225 or 290-300 or 370-375 when the anthranilate synthase

amino acid sequence is aligned with a monomeric *Agrobacterium tumefaciens* anthranilate synthase having SEQ ID NO:4.

62. The transgenic plant of claim 60, wherein the mutation is in the tryptophan-binding
5 pocket.
63. The transgenic plant of claim 60, wherein the mutation is:
(a) at about position 48, replace Val with Phe;
(b) at about position 48, replace Val with Tyr;
10 (c) at about position 51, replace Ser with Phe;
(d) at about position 51, replace Ser with Cys;
(e) at about position 52, replace Asn with Phe;
(f) at about position 293, replace Pro with Ala;
(g) at about position 293, replace Pro with Gly; or
15 (h) at about position 298, replace Phe with Trp; and
wherein the position of the mutation is determined by alignment of the amino acid
sequence of the anthranilate synthase with an *Agrobacterium tumefaciens*
anthranilate synthase amino acid sequence.
- 20 64. The transgenic plant of claim 60, wherein the anthranilate synthase comprises any
one of SEQ ID NO:58-65, 69 or 70.
65. The transgenic plant of claim 63, wherein the *Agrobacterium tumefaciens*
anthranilate synthase amino acid sequence is SEQ ID NO:4.
25
66. The transgenic plant of claim 52 or 56, wherein the isolated DNA further comprises
a plastid transit peptide.

67. The transgenic plant of claim 66, wherein the plastid transit peptide comprises SEQ ID NO:72 or 74.
68. The transgenic plant of claim 52 or 56, wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.
5
69. The transgenic plant of claim 68, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of said plant.
- 10 70. The transgenic plant of claim 69, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.
71. The transgenic plant of claim 52 or 56, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.
15
72. The transgenic plant of claim 52 or 56, wherein the plant is a dicot.
73. The transgenic plant of claim 72, wherein the plant is soybean or canola.
20
74. The transgenic plant of claim 52 or 56, wherein the plant is a monocot.
75. The transgenic plant of claim 72, wherein the plant is maize, rice, wheat, barley or sorghum.
25
76. A seed of the transgenic plant of claim 52 or 56.

77. A transgenic plant comprising an isolated DNA encoding an α domain of anthranilate synthase from *Zea mays* that comprises SEQ ID NO:5 or SEQ ID NO:66 operably linked to a promoter.
- 5 78. The transgenic plant of claim 77, wherein the isolated DNA comprises SEQ ID NO:2, SEQ ID NO:67 or SEQ ID NO:68 operably linked to a promoter.
79. The transgenic plant of claim 77, wherein the α domain of monomeric anthranilate synthase is expressed so as to elevate the level of L-tryptophan in said plant.
- 10 80. The transgenic plant of claim 77, wherein the domain has at least one mutation that increases anthranilate synthase activity or reduces the sensitivity of the domain to inhibition by tryptophan or an analog thereof.
- 15 81. The transgenic plant of claim 77, wherein the mutation is in a tryptophan-binding pocket.
82. The transgenic plant of claim 77, wherein the isolated DNA further encodes a plastid transit peptide.
- 20 83. The transgenic plant of claim 82, wherein the plastid transit peptide comprises SEQ ID NO:72 or 74.
84. The transgenic plant of claim 77, wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.
- 25 85. The transgenic plant of claim 84, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of said plant.

86. The transgenic plant of claim 85, wherein the herbicide resistance comprises
resistance to glyphosate, glufosinate or dalapon.
- 5 87. The transgenic plant of claim 77, wherein the isolated DNA further encodes a
Bacillus thuringiensis protein that, when expressed in a plant, imparts insect
resistance to the plant.
- 10 88. The transgenic plant of claim 77, wherein the plant is a dicot.
- 15 89. The transgenic plant of claim 88, wherein the plant is soybean or canola.
90. The transgenic plant of claim 77, wherein the plant is a monocot.
- 15 91. The transgenic plant of claim 90, wherein the plant is maize, rice, wheat, barley or
sorghum.
92. A seed of the transgenic plant of claim 77.
- 20 93. A method for altering the tryptophan content in a plant comprising:
 (b) introducing into regenerable cells of a plant a transgene comprising an
 isolated DNA encoding a monomeric anthranilate synthase comprising an
 anthranilate synthase α domain and a anthranilate synthase β domain,
 wherein the isolated DNA is operably linked to a promoter functional in a
 plant cell, to yield transformed plant cells; and
 (c) regenerating a plant from said transformed plant cells wherein the cells of
 the plant express the monomeric anthranilate synthase encoded by the
 isolated DNA in an amount effective to increase the tryptophan content in
- 25

the plant relative to the tryptophan content in an untransformed plant of the same genetic background.

94. The method of claim 93, wherein the monomeric anthranilate synthase is an
5 *Agrobacterium tumefaciens*, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc sp.* PCC7120, *Azospirillum brasilense* or *Anabaena M22983* anthranilate synthase.
95. The method of claim 93, wherein the monomeric anthranilate synthase comprises
10 any one of SEQ ID NO:4, 7, 43, 58, 59, 60, 61, 62, 63, 64, 65, 69, 70, 77, 78, 79, 80, 81 or 82.
96. The method of claim 93, wherein the isolated DNA comprises any one of SEQ ID
15 NO:1, 75, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92 or 93.
97. The method of claim 93, wherein the isolated DNA encodes a chimeric monomeric
20 anthranilate synthase comprising a fusion of an anthranilate synthase α domain from one species and an anthranilate synthase β domain from a second species.
98. The method of claim 93 wherein DNA encoding the α domain or the β domain is
25 obtained from *Agrobacterium tumefaciens*, *Anabaena M22983*, *Arabidopsis thaliana*, *Azospirillum brasilense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc sp.* PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco or *Zea mays*.

99. The method of claim 97, wherein the α domain or the β domain is at least a portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45, 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81 , 82, 99, 100, 101, 102 or 103.
- 5 100. The method of claim 93, wherein the isolated DNA further encodes a plastid transit peptide.
101. The method of claim 100, wherein the plastid transit peptide comprises SEQ ID NO:72 or 74.
- 10 102. The method of claim 93, wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.
- 15 103. The method of claim 102, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of said plant.
104. The method of claim 103, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.
- 20 105. The method of claim 93, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.
- 25 106. The method of claim 93, wherein the anthranilate synthase comprises a mutation that increases anthranilate synthase activity or that reduces the sensitivity of the anthranilate synthase to inhibition by tryptophan or an analog thereof.

107. The method of claim 106, wherein the mutation is within amino acid positions 25-60 or 200-225 or 290-300 or 370-375 when the anthranilate synthase amino acid sequence is aligned with a monomeric *Agrobacterium tumefaciens* anthranilate synthase having SEQ ID NO:4.

5

108. The method of claim 106, wherein the mutation is in the tryptophan-binding pocket.

109. The method of claim 106, wherein the mutation is:

- 10 (a) at about position 48, replace Val with Phe;
(b) at about position 48, replace Val with Tyr;
(c) at about position 51, replace Ser with Phe;
(d) at about position 51, replace Ser with Cys;
(e) at about position 52, replace Asn with Phe;
15 (f) at about position 293, replace Pro with Ala;
(g) at about position 293, replace Pro with Gly; or
(h) at about position 298, replace Phe with Trp; and
wherein the position of the mutation is determined by alignment of the amino acid sequence of the anthranilate synthase with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence.

20

110. The method of claim 109 wherein the anthranilate synthase comprises any one of SEQ ID NO:58-65, 69 or 70.

25 111. The method of claim 109 wherein the *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence is SEQ ID NO:4.

112. The method of claim 93, wherein the plant is a dicot.

113. The method of claim 112, wherein the plant is soybean or canola.

114. The method of claim 93, wherein the plant is a monocot.

5

115. The method of claim 114 wherein the plant is maize, rice, wheat, barley or sorghum.

116. A method for altering the tryptophan content in a plant comprising:

10 (a) introducing into regenerable cells of a plant a transgene comprising an isolated DNA encoding an α domain of anthranilate synthase from *Zea mays* that comprises SEQ ID NO:5 or SEQ ID NO:66, operably linked to a promoter functional in a plant cell and to yield transformed plant cells; and

15 (b) regenerating a plant from said transformed plant cells wherein the cells of the plant express the anthranilate synthase encoded by the isolated DNA in an amount effective to increase the tryptophan content in the plant relative to the tryptophan content in an untransformed plant of the same genetic background.

20 117. The method of claim 116, wherein the α domain of anthranilate synthase has a mutation that increases anthranilate synthase activity or reduces the sensitivity of the domain to inhibition by tryptophan or an analog thereof.

25 118. The method of claim 116 wherein the mutation is in a tryptophan-binding pocket.

119. The method of claim 116, wherein the plant is a dicot.

120. The method of claim 119, wherein the plant is soybean or canola.

121. The method of claim 116, wherein the plant is a monocot.

5 122. The method of claim 121, wherein the plant is maize, rice, wheat, barley or sorghum.

123. The method of claim 116, wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.

10 124. The method of claim 123, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of said plant.

15 125. The method of claim 124, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.

126. The method of claim 116, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.

20 127. A method for making an animal feed or a human food comprising:
(a) introducing into regenerable cells of a plant a transgene comprising an isolated DNA encoding a monomeric anthranilate synthase comprising an anthranilate synthase α domain and a anthranilate synthase β domain, wherein the isolated DNA is operably linked to a promoter functional in a plant cell, to yield transformed plant cells; and
25 (b) regenerating a plant from said transformed plant cells wherein the cells of the plant express the monomeric anthranilate synthase encoded by the

isolated DNA in an amount effective to increase the tryptophan content in the plant relative to the tryptophan content in an untransformed plant of the same genetic background.

- 5 128. The method of claim 127, wherein the monomeric anthranilate synthase is an *Agrobacterium tumefaciens*, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc* sp. PCC7120, *Azospirillum brasiliense* or *Anabaena* M22983 anthranilate synthase.
- 10 129. The method of claim 127, wherein the monomeric anthranilate synthase comprises any one of SEQ ID NO:4, 7, 43, 58, 59, 60, 61, 62, 63, 64, 65, 69, 70, 77, 78, 79, 80, 81 or 82.
- 15 130. The method of claim 127, wherein the isolated DNA comprises any one of SEQ ID NO:1, 75, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92 or 93.
- 20 131. The method of claim 127, wherein the isolated DNA encodes a chimeric monomeric anthranilate synthase comprising a fusion of an anthranilate synthase α domain from one species and an anthranilate synthase β domain from a second species.
- 25 132. The method of claim 127, wherein DNA encoding the α domain or the β domain is obtained from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco or *Zea mays*.

133. The method of claim 127, wherein the α domain or the β domain is at least a portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45, 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81 , 82, 99, 100, 101, 102 or
5 103.
134. The method of claim 127, wherein the anthranilate synthase comprises a mutation that increases anthranilate synthase activity or reduces the sensitivity of the anthranilate synthase to inhibition by tryptophan or an analog thereof.
10
135. The method of claim 134, wherein the mutation is in the tryptophan-binding pocket.
136. The method of claim 134, wherein the mutation is:
15 (a) at about position 48, replace Val with Phe;
(b) at about position 48, replace Val with Tyr;
(c) at about position 51, replace Ser with Phe;
(d) at about position 51, replace Ser with Cys;
(e) at about position 52, replace Asn with Phe;
20 (f) at about position 293, replace Pro with Ala;
(g) at about position 293, replace Pro with Gly; or
(h) at about position 298, replace Phe with Trp; and
wherein the position of the mutation is determined by alignment of the amino acid sequence of the anthranilate synthase with an *Agrobacterium tumefaciens*
25 anthranilate synthase amino acid sequence.
137. The method of claim 136, wherein the *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence comprises SEQ ID NO:4.

138. The method of claim 134, wherein the anthranilate synthase comprises any one of SEQ ID NO:58-65, 69 or 70.

5 139. The method of claim 127, wherein the isolated DNA further encodes a plastid transit peptide.

140. The method of claim 139, wherein the plastid transit peptide comprises SEQ ID NO:72 or 74.

10 141. The method of claim 127 wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.

142. The method of claim 141, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of said plant.

143. The method of claim 142, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.

20 144. The method of claim 127, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.

145. The method of claim 127, wherein the plant is a dicot.

25 146. The method of claim 145 wherein the plant is soybean or canola.

147. The method of claim 127 wherein the plant is a monocot.

148. The method of claim 147 wherein the plant is maize, rice, wheat, barley or sorghum.
- 5 149. An animal feed or human food comprising at least a portion of a plant that comprises an isolated DNA encoding a monomeric anthranilate synthase comprising an anthranilate synthase α domain and a anthranilate synthase β domain, wherein the cells of the plant can express the monomeric anthranilate synthase encoded by the isolated DNA.
- 10 150. The animal feed or human food of claim 149, wherein the cells of the plant can express the monomeric anthranilate synthase in an amount effective to increase the tryptophan content in the plant relative to the tryptophan content in an untransformed plant of the same genetic background.
- 15 151. The animal feed or human food of claim 149, wherein the portion of the plant comprises a seed, a leaf, a stem, a root, a tuber, or a fruit.
152. The animal feed or human food of claim 149, wherein the monomeric anthranilate synthase is an *Agrobacterium tumefaciens*, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella melitensis*, *Nostoc sp. PCC7120*, *Azospirillum brasilense* or *Anabaena M22983* anthranilate synthase.
- 20 153. The animal feed or human food of claim 149, wherein the monomeric anthranilate synthase comprises any one of SEQ ID NO:4, 7, 43, 58, 59, 60, 61, 62, 63, 64, 65, 69, 70, 77, 78, 79, 80, 81 or 82.

154. The animal feed or human food of claim 149, wherein the isolated DNA comprises any one of SEQ ID NO:1, 75, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92 or 93.
- 5 155. The animal feed or human food of claim 149, wherein the isolated DNA encodes a chimeric monomeric anthranilate synthase comprising a fusion of an anthranilate synthase α domain from one species and an anthranilate synthase β domain from a second species.
- 10 156. The animal feed or human food of claim 155, wherein DNA encoding the α domain or the β domain is obtained from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis thaliana*, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena gracilis*, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia marcescens*, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco or *Zea mays*.
- 15 157. The animal feed or human food of claim 149, wherein the α domain or the β domain is at least a portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45, 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81 , 82, 99, 100, 101, 102 or 103.
- 20 158. The animal feed or human food of claim 149, wherein the anthranilate synthase comprises a mutation that increases anthranilate synthase activity or reduces the sensitivity of the anthranilate synthase to inhibition by tryptophan or an analog thereof.
- 25

159. The animal feed or human food of claim 158, wherein the mutation is within amino acid positions 25-60 or 200-225 or 290-300 or 370-375 when the anthranilate synthase amino acid sequence is aligned with a monomeric *Agrobacterium tumefaciens* anthranilate synthase having SEQ ID NO:4.

5

160. The animal feed or human food of claim 158, wherein the mutation is in a tryptophan-binding pocket.

161. The animal feed or human food of claim 158, wherein the mutation is:

- 10 (a) at about position 48, replace Val with Phe;
 (b) at about position 48, replace Val with Tyr;
 (c) at about position 51, replace Ser with Phe;
 (d) at about position 51, replace Ser with Cys;
 (e) at about position 52, replace Asn with Phe;
15 (f) at about position 293, replace Pro with Ala;
 (g) at about position 293, replace Pro with Gly; or
 (h) at about position 298, replace Phe with Trp; and
 wherein the position of the mutation is determined by alignment of the amino acid sequence of the anthranilate synthase with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence.

20

162. The animal feed or human food of claim 149, wherein the anthranilate synthase comprises any one of SEQ ID NO:58-65, 69 or 70.

25

163. The animal feed or human food of claim 161, wherein the *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence is SEQ ID NO:4.

164. The animal feed or human food of claim 149, wherein the isolated DNA further
encodes a plastid transit peptide.
165. The animal feed or human food of claim 164, wherein the plastid transit peptide
5 comprises SEQ ID NO:72 or 74.
166. The animal feed or human food of claim 149, wherein the isolated DNA further
encodes a selectable marker gene or a reporter gene.
- 10 167. The animal feed or human food of claim 166, wherein the selectable marker gene,
. when expressed in a plant, imparts herbicide resistance to cells of said plant.
168. The animal feed or human food of claim 167, wherein the herbicide resistance
comprises resistance to glyphosate, glufosinate or dalapon.
- 15 169. The animal feed or human food of claim 149, wherein the isolated DNA further
encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts
insect resistance to the plant.
- 20 170. The animal feed or human food of claim 149, wherein the plant is a dicot.
171. The animal feed or human food of claim 170, wherein the plant is soybean or
canola.
- 25 172. The animal feed or human food of claim 149, wherein the plant is a monocot.
173. The animal feed or human food of claim 172, wherein the plant is maize, rice,
wheat, barley or sorghum.

174. An isolated DNA encoding an anthranilate synthase comprising a polypeptide having at least 90% sequence identity with SEQ ID NO:4.
- 5 175. An isolated DNA encoding an anthranilate synthase, comprising a DNA having at least 60% sequence identity with SEQ ID NO:1.
- 10 176. The isolated DNA of claim 174 or 175, wherein the isolated DNA comprises at least twenty nucleotides and that hybridizes to the complement of a DNA having SEQ ID NO:1 under stringent hybridization conditions.
177. The isolated DNA of claim 176, wherein the stringent hybridization conditions comprise washing at 42°C in 0.2 x SSC.
- 15 178. The isolated DNA of claim 176, wherein the isolated DNA comprises any one of SEQ ID NO:9-42 or 46-56.
179. An isolated DNA encoding an α domain of anthranilate synthase from *Zea mays*, that comprises amino acid sequence SEQ ID NO:5 or SEQ ID NO:66.
- 20 180. An isolated DNA encoding an α domain of anthranilate synthase from *Zea mays* that comprises nucleotide sequence SEQ ID NO:2, SEQ ID NO:67 or SEQ ID NO:68.
- 25 181. The isolated DNA of claim 179 or 180, wherein the α domain of anthranilate synthase can be expressed in a plant so as to elevate the level of L-tryptophan in the plant.

182. The isolated DNA of claim 179 or 180, wherein the domain has at least one mutation that reduces the sensitivity of the domain to inhibition by tryptophan or an analog thereof.
- 5 183. The isolated DNA of claim 179 or 180, wherein the mutation is in a tryptophan-binding pocket.
184. The isolated DNA of claim 179 or 180, wherein the isolated DNA further encodes a selectable marker gene or a reporter gene.
- 10 185. The isolated DNA of claim 184, wherein the selectable marker gene, when expressed in a plant, imparts herbicide resistance to cells of a plant.
- 15 186. The isolated DNA of claim 185, wherein the herbicide resistance comprises resistance to glyphosate, glufosinate or dalapon.
187. The isolated DNA of claim 179 or 180, wherein the isolated DNA further encodes a *Bacillus thuringiensis* protein that, when expressed in a plant, imparts insect resistance to the plant.
- 20 188. The isolated DNA of claim 179 or 180, wherein the plant is a dicot.
189. The isolated DNA of claim 188, wherein the plant is soybean or canola.
- 25 190. The isolated DNA of claim 179 or 180, wherein the plant is a monocot.
191. The isolated DNA of claim 190, wherein the plant is maize, rice, wheat, barley or sorghum.

192. The isolated DNA of claim 179 or 180, wherein the isolated DNA encoding the anthranilate synthase comprises a promoter operably linked thereto.

5 193. A vector comprising the isolated DNA of claim 179 or 180.

194. An isolated DNA encoding a mutant anthranilate synthase, wherein the mutation comprises:

- (a) at about position 48, replace Val with Phe;
- (b) at about position 48, replace Val with Tyr;
- (c) at about position 51, replace Ser with Phe;
- (d) at about position 51, replace Ser with Cys;
- (e) at about position 52, replace Asn with Phe;
- (f) at about position 293, replace Pro with Ala;
- (g) at about position 293, replace Pro with Gly; or
- (h) at about position 298, replace Phe with Trp; and

10 wherein the position of the mutation is determined by alignment of the amino acid sequence of the anthranilate synthase with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence.

20

195. The isolated DNA of claim 194, wherein the anthranilate synthase comprises any one of SEQ ID NO:58-65.

25

196. The isolated DNA of claim 194, wherein the *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence is SEQ ID NO:4.

197. A method for producing tryptophan comprising: culturing a prokaryotic or eukaryotic host cell comprising an isolated DNA under conditions sufficient to

express a monomeric anthranilate synthase encoded by the isolated DNA, wherein
the monomeric anthranilate synthase comprises an anthranilate synthase α domain
and a anthranilate synthase β domain, and wherein the conditions sufficient to
express a monomeric anthranilate synthase comprise nutrients and precursors
5 sufficient for the host cell to synthesize tryptophan utilizing the monomeric
anthranilate synthase.

198. The method of claim 197, wherein the method further comprises producing a
phenylpropanoid, a flavonoid, an isoflavonoid, an indole, an indole alkaloid, or an
10 indole glucosinolate.
199. The method of claim 197, wherein the monomeric anthranilate synthase is an
Agrobacterium tumefaciens, *Rhizobium meliloti*, *Mesorhizobium loti*, *Brucella*
melitensis, *Nostoc sp.* PCC7120, *Azospirillum brasiliense* or *Anabaena* M22983
15 anthranilate synthase.
200. The method of claim 197, wherein the monomeric anthranilate synthase
comprises any one of SEQ ID NO:4, 7, 43, 58, 59, 60, 61, 62, 63, 64, 65, 69, 70,
77, 78, 79, 80, 81 or 82.
20
201. The method of claim 197, wherein the isolated DNA comprises any one of SEQ
ID NO:1, 75, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92 or 93.
202. The method of claim 197, wherein the isolated DNA encodes a chimeric
25 monomeric anthranilate synthase comprising a fusion of an anthranilate synthase
 α domain from one species and an anthranilate synthase β domain from a second
species.

203. The method of claim 197, wherein DNA encoding the α domain or the β domain
is obtained from *Agrobacterium tumefaciens*, *Anabaena* M22983, *Arabidopsis*
thaliana, *Azospirillum brasiliense*, *Brucella melitensis*, *Escherichia coli*, *Euglena*
gracilis, *Mesorhizobium loti*, *Nostoc* sp. PCC7120, *Rhizobium meliloti*, *Ruta*
5 *graveolens*, *Rhodopseudomonas palustris*, *Salmonella typhimurium*, *Serratia*
marcescens, *Sulfolobus solfataricus*, soybean, rice, cotton, wheat, tobacco or *Zea*
mays.
204. The method of claim 197, wherein the α domain or the β domain is at least a
portion of any one of amino acid sequences SEQ ID NO:4, 5, 6, 7, 8, 43, 44, 45,
10 58, 59, 60, 61, 62, 63, 64, 65, 66, 69, 70, 77, 78, 79 80, 81 , 82, 99, 100, 101, 102
or 103.
205. The method of claim 197, wherein the anthranilate synthase comprises a mutation
15 that increases anthranilate synthase activity or reduces the sensitivity of the
anthranilate synthase to inhibition by tryptophan or an analog thereof.
206. The method of claim 205, wherein the mutation is within amino acid positions 25-
60 or 200-225 or 290-300 or 370-375 when the anthranilate synthase amino acid
20 sequence is aligned with a monomeric *Agrobacterium tumefaciens* anthranilate
synthase having SEQ ID NO:4.
207. The method of claim 205, wherein the mutation is in a tryptophan-binding pocket.
- 25 208. The method of claim 205, wherein the mutation is:
(a) at about position 48, replace Val with Phe;
(b) at about position 48, replace Val with Tyr;
(c) at about position 51, replace Ser with Phe;

(d) at about position 51, replace Ser with Cys;
(e) at about position 52, replace Asn with Phe;
(f) at about position 293, replace Pro with Ala;
(g) at about position 293, replace Pro with Gly; or
5 (h) at about position 298, replace Phe with Trp; and
wherein the position of the mutation is determined by alignment of the amino acid sequence of the anthranilate synthase with an *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence.

10 209. The method of claim 197, wherein the anthranilate synthase comprises any one of SEQ ID NO:58-65, 69 or 70.

210. The method of claim 208, wherein the *Agrobacterium tumefaciens* anthranilate synthase amino acid sequence is SEQ ID NO:4.

15

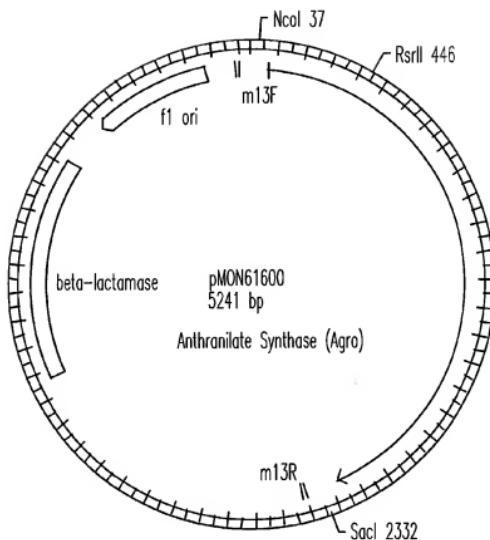
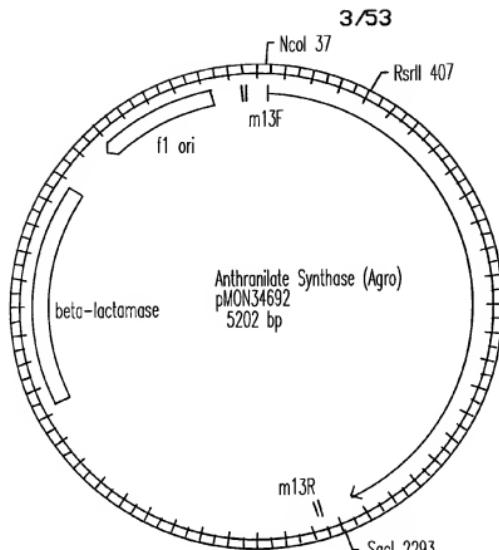
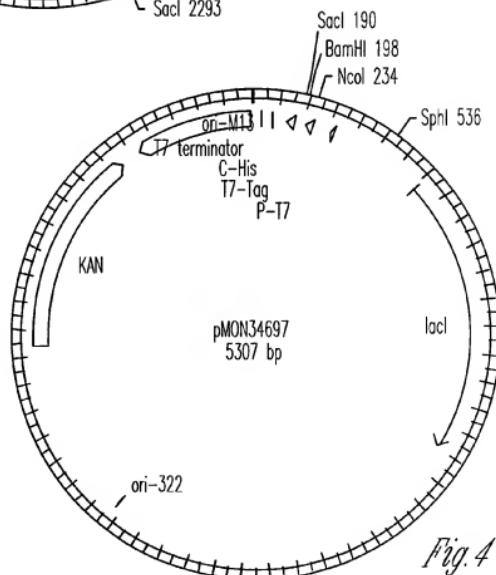


Fig. 1

2/53

1 MVTIIQDDGAETYETKGGIOVSRKRRTDYZANAIIDNYIEKLDSHRGAVFS 50
 | :| :||| . ||| . :| :| ..| :| :||| . |||
 1 MAAVILEDGAESYTTKGGIVVTRRRREASYSDAIAGYVDRLDERRGAVFS 50
 |||||
 51 SNYEYPGRYTRWDTAIVDPLGISCFCGRKMWIEAYNGRGEVLLDFITEKL 100
 |||||
 51 SNYEYPGRYTRWDTAVVDPPLAISSFGRSLWIEAYNERGEVLLALIAEDL 100
 101 KATPDLTGLASSTRRLDLTVNEPDRVFTEEERSKIPTVFTRALRAIVDLFY 150
 | . :| :| . . ||| :||| :||| :||| :||| :||| :||| :|||
 101 KSVADITGLSAAARRLDTINEPDRVFTEEERSKMPVTFLRAVTNLFH 150
 151 SSDSAIGLFGAFGYDLAQFDIAKLSLARPEDQRDMVLFLPDEILVVDH 200
 | . :| :| . . ||| :||| :||| :||| :||| :||| :|||
 151 SEEDSNLGLYGAFFGYDLAQFDIAELKLSLRPDDQRDMVLFLPDEILVVDH 200
 201 YSAKAWIDRYDFEKDGMTTDGKSSDTPDPFKTTDTIPPKGDHRPGEYSE 250
 | . . ||| :||| :||| :||| :||| :||| :||| :||| :|||
 201 YAAKAWIDRYDFARENLISTEGKAADIAPFPRSVDSIPPNGDHRPGEYAE 250
 251 LVVKAKESFRRGDLFEVVPQGKFMERCESNPSAISRRLKAINPSPYSFFI 300
 251 LVVKAKESFRRGDLFEVVPQGKFYERCESRPSEISNRLKAINPSPYSFFI 300
 301 NLGDQEYLVGASPEMFVRVSGRRRIETCPISGTIKRGDDPIADSEQILKLL 350
 | . . ||| :||| :||| :||| :||| :||| :||| :|||
 301 NLGNQEYLVGASPEMFVRVSGRRRIETCPISGTIKRGDDPIADSEQILKLL 350
 351 NSKKDESELTCMCSDVDRNDKSRVCCEPGSVKVIGRRQIEMYSRLIHTVDHI 400
 | . . ||| :||| :||| :||| :||| :||| :|||
 351 NSKKDESELTCMCSDVDRNDKSRVCCEPGSVKVIGRRQIEMYSRLIHTVDHI 400
 401 EGRLRDDMDAFDGFSLSHAWAVTGTGAPKWLWAMRFIEGHEKSPPRAWYGGAI 450
 | . . ||| :||| :||| :||| :||| :||| :|||
 401 EGRLRDDMDAFDGFSLSHAWAVTGTGAPKWLWAMRFIESHEKSPPRAWYGGAI 450
 451 GMVGFNGDMNTGLTLRTIRIKDIAEVRAVAGATLLNDSNPQEEEAETELKA 500
 | . . ||| :||| :||| :||| :||| :||| :|||
 451 GMVGFNGDMNTGLTLRTIRIKDIAEVRAVAGATLLYDSNPQEEEAETELKA 500
 501 SAMISAIRDAKGTNSAATKRAAKVGTGVKILLVDHEDSFVHTLANYFRQ 550
 | . . ||| :||| :||| :||| :||| :||| :|||
 501 SAMIAAIRDAKSANSAKSARDVAAVGAGVSILLVDHEDSFVHTLANYFRQ 550
 551 TGATVSTVRSPVAAADVFDRFQPDFLVVLSPGPGSTDFDCATIKAARARD 600
 | . . ||| :||| :||| :||| :||| :||| :|||
 551 TGASVTTVRTPVAAEEIFDRVKPDLVVLSPGPGBTPKDFDCATIKKARARD 600
 601 LPIFGVCLGLQALAEAYGGELRQLAVPMHGPSPSIRVLEPGLVFSGLGKE 650
 | . . ||| :||| :||| :||| :||| :||| :|||
 601 LPIFGVCLGLQALAEAYGGDLRQLAIPMHGPSPSIRVLEPGLVFSGLGKE 650
 651 VTVGRYHSIFADPATLERPDFITAESEDGTIMGIEHAKEPVAAVQFHPES 700
 | . . ||| :||| :||| :||| :||| :|||
 651 VTVGRYHSIFADPSNLPREFVITAESDGTIMGIEHSKEPVAAVQFHPES 700
 701 IMTLGQDAGMRMIENVVHLTRKAKTKAA 729
 | . . ||| :||| :||| :|||
 701 IMTLGGDAGMRMIENVVAAHLAKRAKTKAA 729

Fig.2

*Fig. 3**Fig. 4*

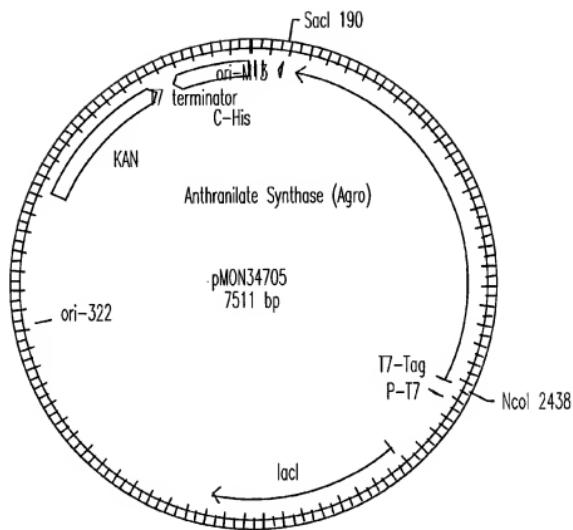


Fig.5

5/53

Agrobacterium_Trpeg Sulfobolus_Trpe	MVTTIQQDDGAETYETKGGIQVSRSRKERTDYZANADNYIEKLDSHRGAVFESSNYEYPGRYT -----MEVHTSEFAASPEFEVKCERDFKVAGIES-----
Agrobacterium_Trpeg Sulfobolus_Trpe	R-WDTAIVDPPIGISCFGRKMWEAYNGREVLIDFTIEKIKATPDLTIGASSTRRLDT -----IGGPQTKARYSVIASTWNG-----YKIHDPP-----VNIL
Agrobacterium_Trpeg Sulfobolus_Trpe	VNEBPDVRVTEBRSSKIPVTALRIVDFYSSADSAAIGLFGAFGFGYDIAFQFDALKLSLA NG-----YLKDLKLADIPGIFKG-----GNIGYISYDAVRFWKEIRDLKP
Agrobacterium_Trpeg Sulfobolus_Trpe	RPEPDORDMVLFLPDEILVVYDHYSAKAWIDRYDFERDGMTIDGKSSDITDPDFKTTDTIPP AEEDWPYAEFTPDNIIYDHNEGKYVNV-----ADLSSVGGCGD1GEFKVSYTDESIN
Agrobacterium_Trpeg Sulfobolus_Trpe	XGDHRPGEYSLIVVKAESFRRGDLIFEVYPGQKEMERCEMSNPSAISRBLKAINPSPYSEF K--N--S-YERIVSESLRTYRSGYLFQVVLSRTRYIIFSGDPLRITYNRRINSPYMFY
Agrobacterium_Trpeg Sulfobolus_Trpe	TNLGDQEYLVAGASPEMFVRSGRRIETCPISGTIKRGDDPLADSEQTIKLNSKDESEL LKF-DEKYLIGSSPELLPRVQDNIVETYPIAGTRPRGAQEQEDLKLIEELMNSERKDAEH
Agrobacterium_Trpeg Sulfobolus_Trpe	TMCSDVDNRNDGSRVCEPGSYKVKIGRROLEMYSRJLHTVDIEGRLRDMDAFDGTLISHAW IMLVDLARNLIGKVCVPGTVKPELMYVEKYSHVQHIVSKVIGTLKKYNALNVLSATFF

Fig. 6A

6/53

Agrobacterium TrpEG Sulfolobus TrpE	AVVYTGAPKLIWAMRFIEGHKSPPRANYGGAGIMVGNGDMNTGILTIRTRIKDGIAEVRA AGTVSGRPKPAMNI IETLEYEKRGPYAGAVFTISADGNAEFAIAIRTAFLNKELLRIHA
Agrobacterium TrpEG Sulfolobus TrpE	GATLLINDSNPQEPEAETELMASAMISAIRDAGTNSAAKRDAAKVGETVKILLVDHEDS GAGIVYDSNPESEYFETEHKKALKTAIGVR-----
Agrobacterium TrpEG Sulfolobus TrpE	FVHILANYFRQTFATVSTVSPVAADFDRFQPDVLVLSPGPGSPTDFOCKATIKAARAR
Agrobacterium TrpEG Sulfolobus TrpE	DLPIFGVCLGLQIAEAAYGGELRQLAVPMHGKPSRIRVLEPLGVFGSLGKEVTGRYHSI
Agrobacterium TrpEG Sulfolobus TrpE	FADPATLPRDFITIAESEDGTIMGIEHAKEPVAAVQFHPESTINTLGQDAGMRMIENVVVH
Agrobacterium TrpEG Sulfolobus TrpE	LTRKAKTKAA

Fig 6B

V48F-F: CCATCGGGCGGTTTTCGTCAACTATG (SEQ ID NO: 9)
 V48F-R: CATA GTGGACGAAAAAAaCGGCCGCGATGG (SEQ ID NO:10)

V48Y-F: CCATCGGGCGGTTTCGTCCAACCTATGAATATCC (SEQ ID NO:11)
 V48Y-R: GGATATTCATAGTGGACGAAAAAatACGCCGCATGG (SEQ ID NO:12)

V48W-F: CCATCGGGCGGTTTCGTCCAACCTATGAATATCC (SEQ ID NO:13)
 V48W-R: GGATATTCATAGTGGACGAAAAAccAGGCCGCATGG (SEQ ID NO:14)

S50K-F: CCATCGGGCGGTTTaGTCCAACTATGAATATCC (SEQ ID NO:15)
 S50K-R: GGATATTCATAGTGGACTTaaaACCGGCCGCATGG (SEQ ID NO:16)

S51C-F: CGCGGGTTTTTCGTGCAACTATGAATATCCGGG (SEQ ID NO:17)
 S51C-F: CCCGGATTATCATAGTTGcACGAAAAAACCGCGC (SEQ ID NO:18)

S51F-F: CGCGGGTTTTTCGTGCAACTATGAATATCCGGG (SEQ ID NO:19)
 S51F-R: GCCCGGATATTCA TAGTTGcACGAAAAAACCGCG (SEQ ID NO:20)

S51I-F: CGCGGGTTTTTCGatCAACTATGAATATCCGGG (SEQ ID NO:21)
 S51I-R: GCCCGGATATTCA TAGTTGatCGAAAAAAACCGCG (SEQ ID NO:22)

S51L-F: GGC GGTTTTTCGTCAACTATGAATATCCGGG (SEQ ID NO:23)
 S51L-R: GGC GGATATTCA TAGTTGatCGAAAAAAACCGCG (SEQ ID NO:24)

S51M-F: CGGC GGTTTTTCGatgAACTATGAATATCCGGCCG (SEQ ID NO:25)
 S51M-R: CGGCCGGATATTCA TAGTTGatCGAAAAAAACCGGCCG (SEQ ID NO:26) *Pig. 7A*

S51T-F: CGCGGGTTTCGacCCAACATGAATATCCGGGC (SEQ ID NO:27)
 S51T-R: GCCCCGATATTCAAGTGGtCGAAAAAACGGCG (SEQ ID NO:28)

S51V-F: GGCGCGGTTTCGgtCAACTATGAATAATCCGGGC (SEQ ID NO:29)
 S51V-R: GCCCCGATATTCAAGTGTacCGAAAAAACGGGCC (SEQ ID NO:30)

S51Y-F: GCGCGGGTTTCGtaCAACTATGAATATCCGGGC (SEQ ID NO:31)
 S51Y-R: GCCCCGGATATTCAAGTtgtACGAAAAAACGGGCC (SEQ ID NO:32)

N52F-F: CGGGCGGGTTTTTCGtCCTtCTATGAATAATCCGGG (SEQ ID NO:33)
 N52F-R: CCCGGATATTCAAGAaggACGAAAGGACGAAACCGGCCG (SEQ ID NO:34)

P293A-F: CTGAAGGGGATCAACqCGTCGCCCTATTc (SEQ ID NO:35)
 P293A-R: GAATAGGGGAGCGGTGATCGCCCTTCAg (SEQ ID NO:36)

P293G-F: CCTGAAGGGGATCAACqGGTGCGCCCTATTCC (SEQ ID NO:37)
 P293G-R: GGAATAGGGGGAGCCGGTGTGATGCCCTTCAGG (SEQ ID NO:38)

F298A-F: CGTCGCCCTATTCCgcCTTCATCAATCTCCGGCG (SEQ ID NO:39)
 F298A-R: CGCCGAGATGTGATGAAGggCGGAATAGGGCCGACG (SEQ ID NO:40)

F298W-F: CGTCGCCCTATTCCCTgtTTCATCAATCTCCGGCG (SEQ ID NO:41)
 F298W-R: CGCCGAGATGTGATGAACCAGGAATAGGGCCGACG (SEQ ID NO:42)

Hg. 2B

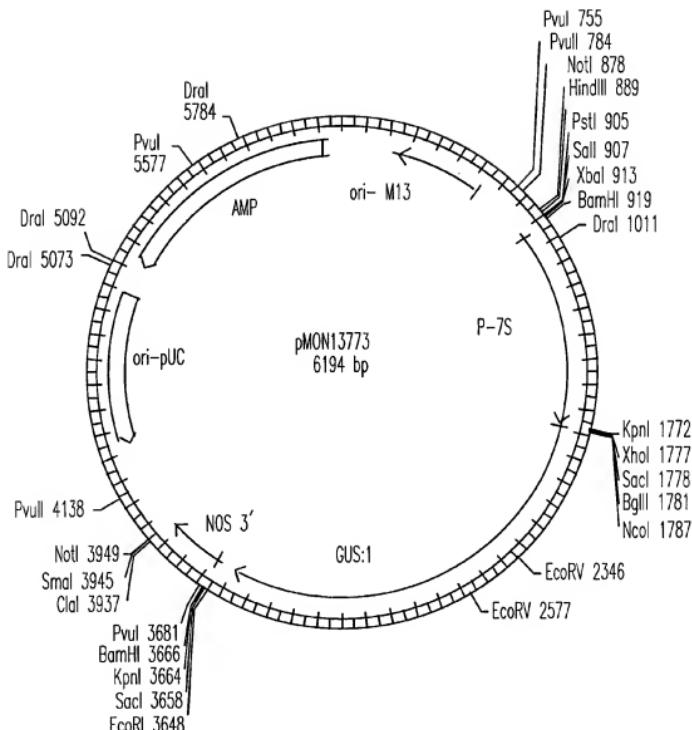


Fig. 8

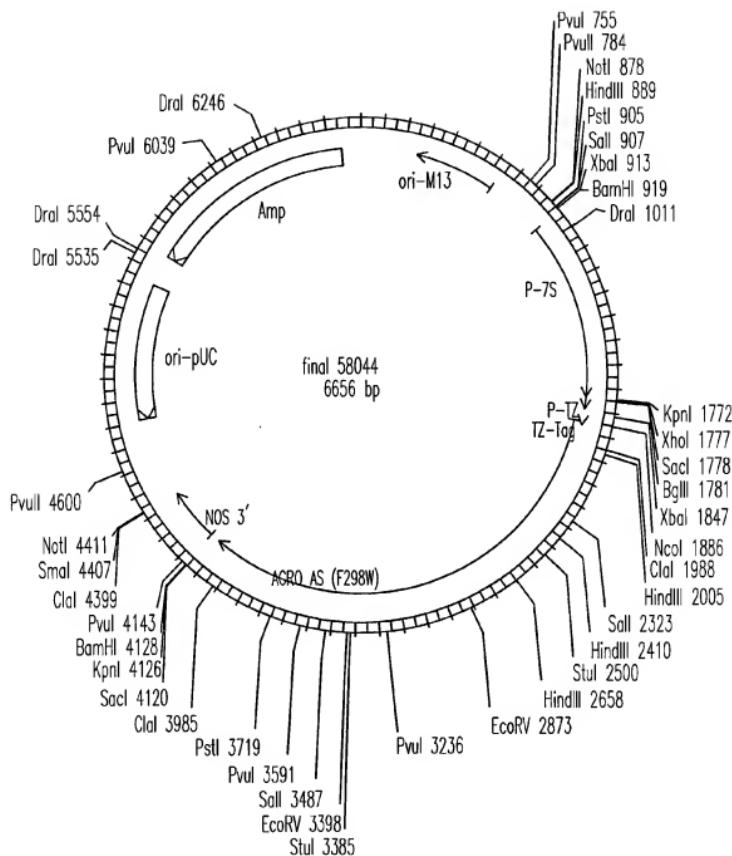


Fig. 9

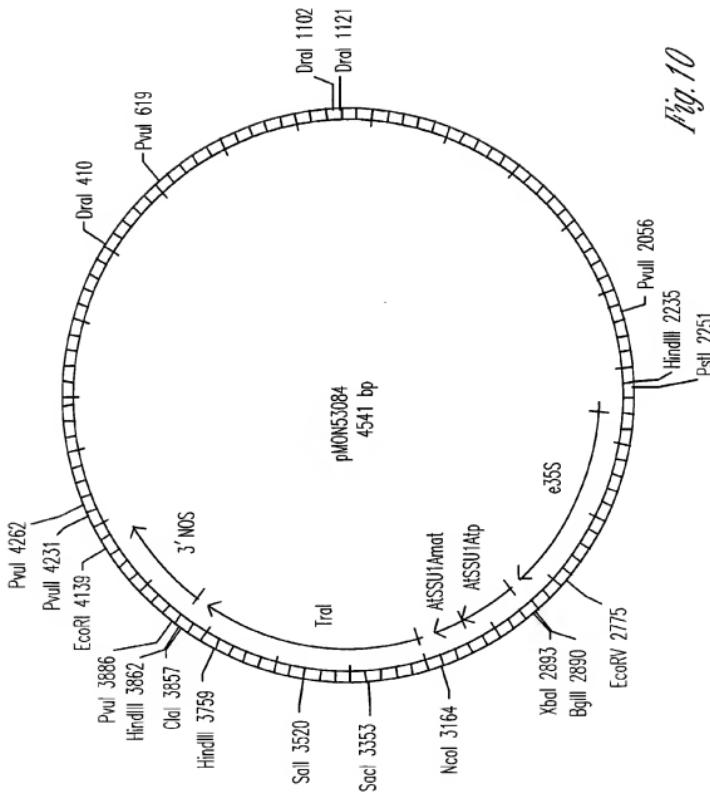


Fig. 10

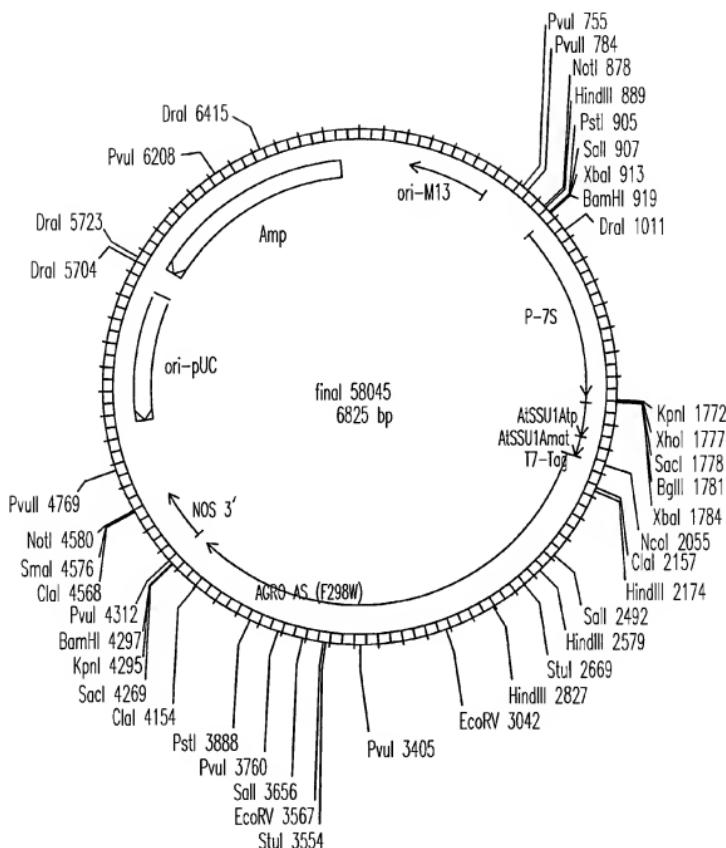


Fig. 11

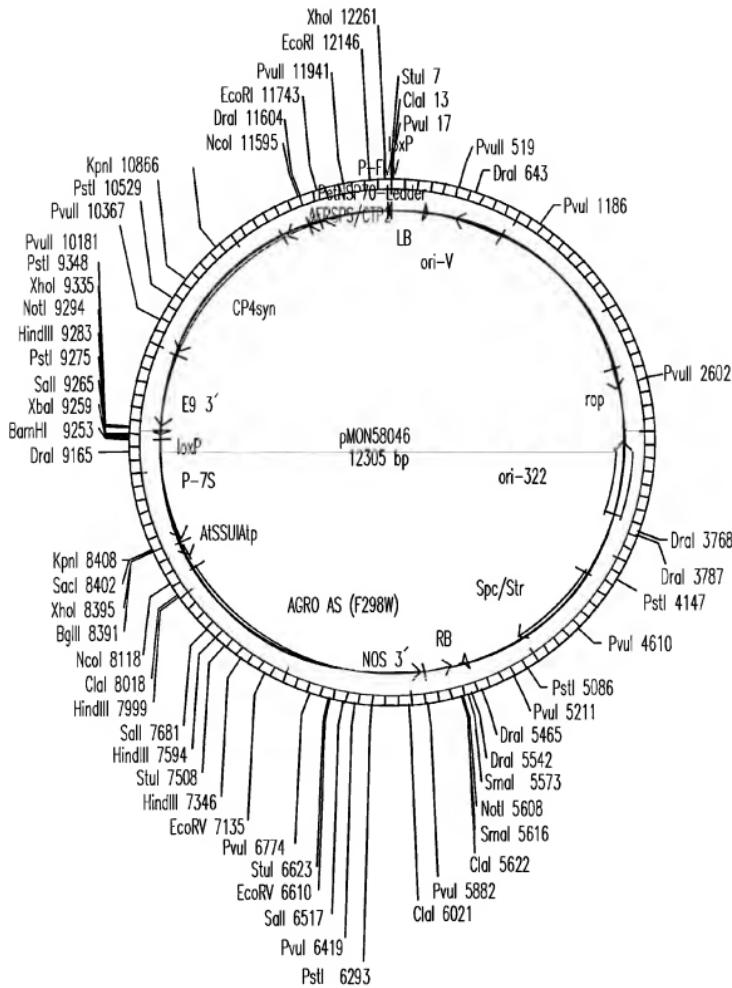


Fig. 12

14/53

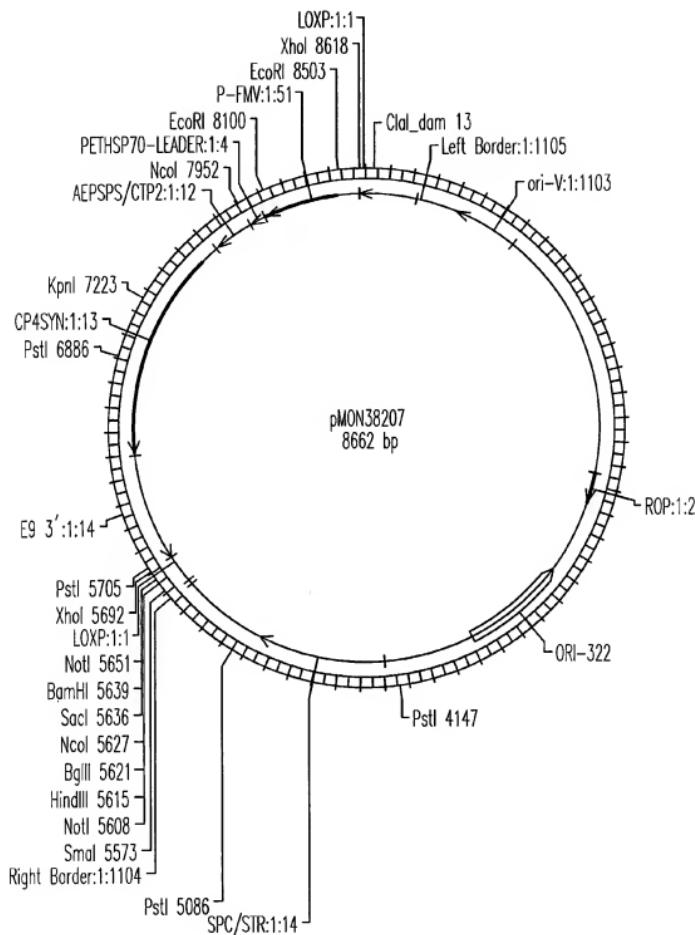


Fig. 13

15/53

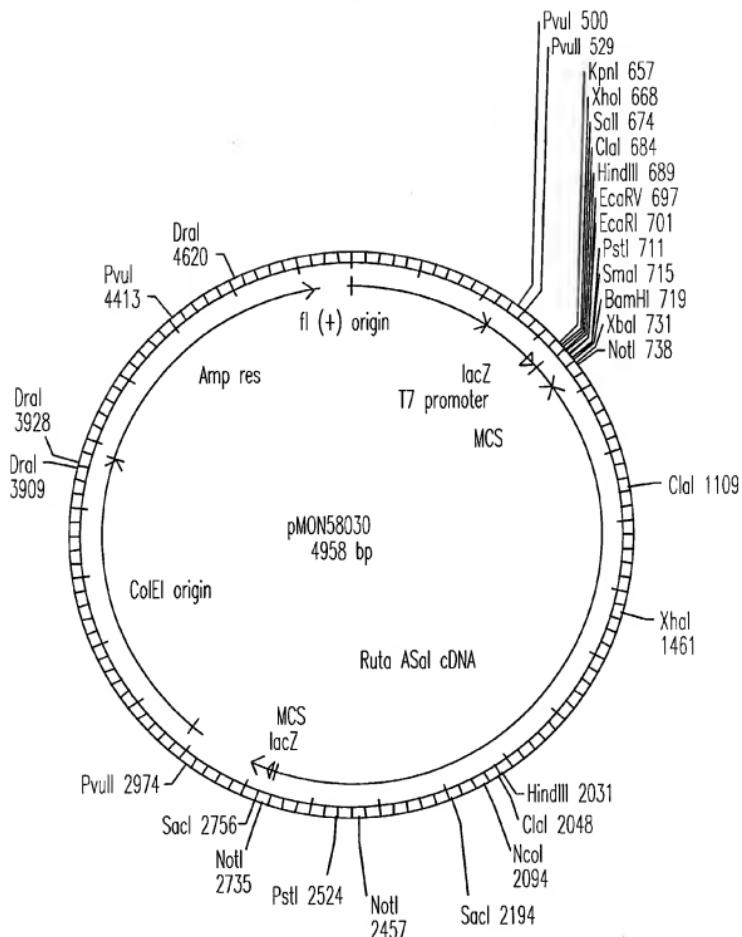


Fig. 14

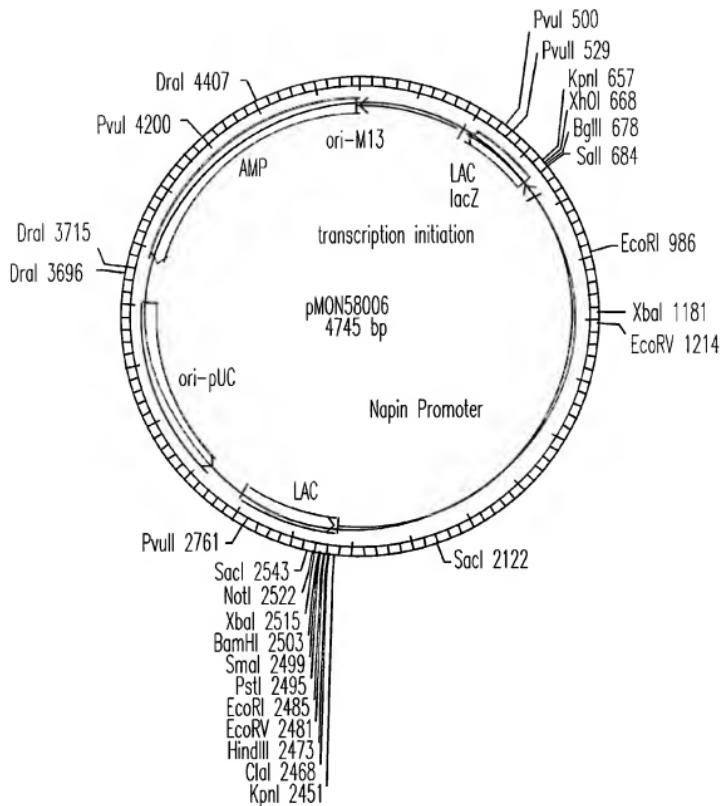


Fig. 15

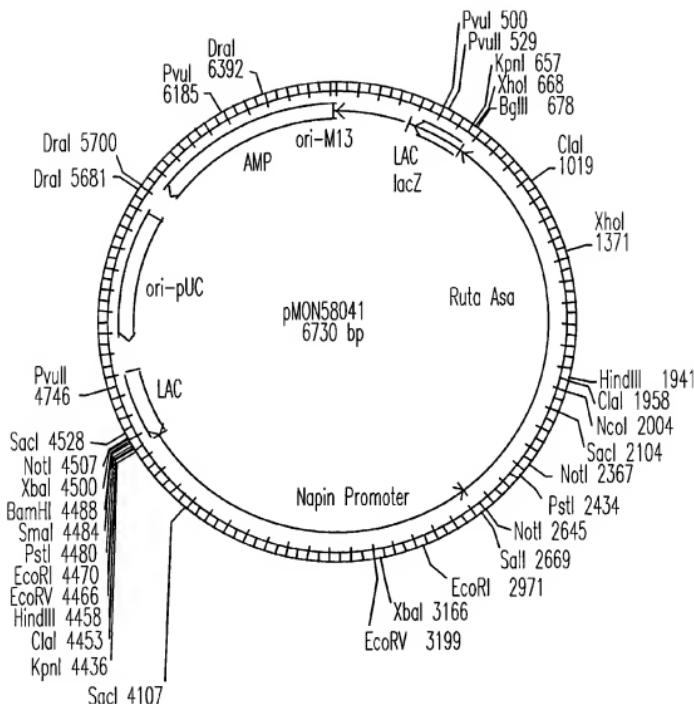


Fig. 16

18/53

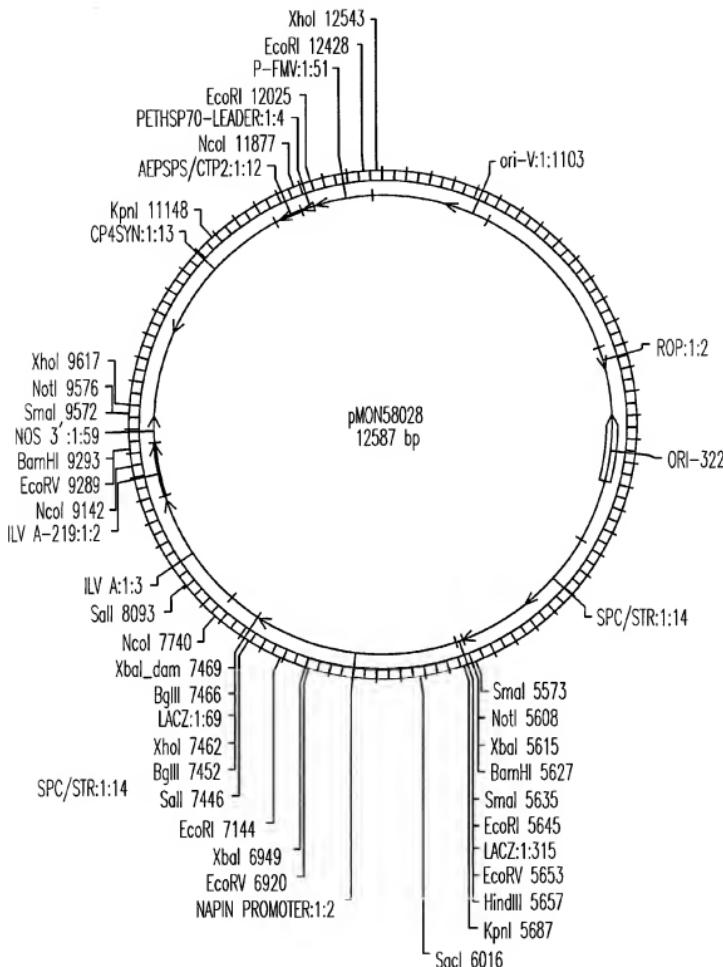


Fig. 17

19/53

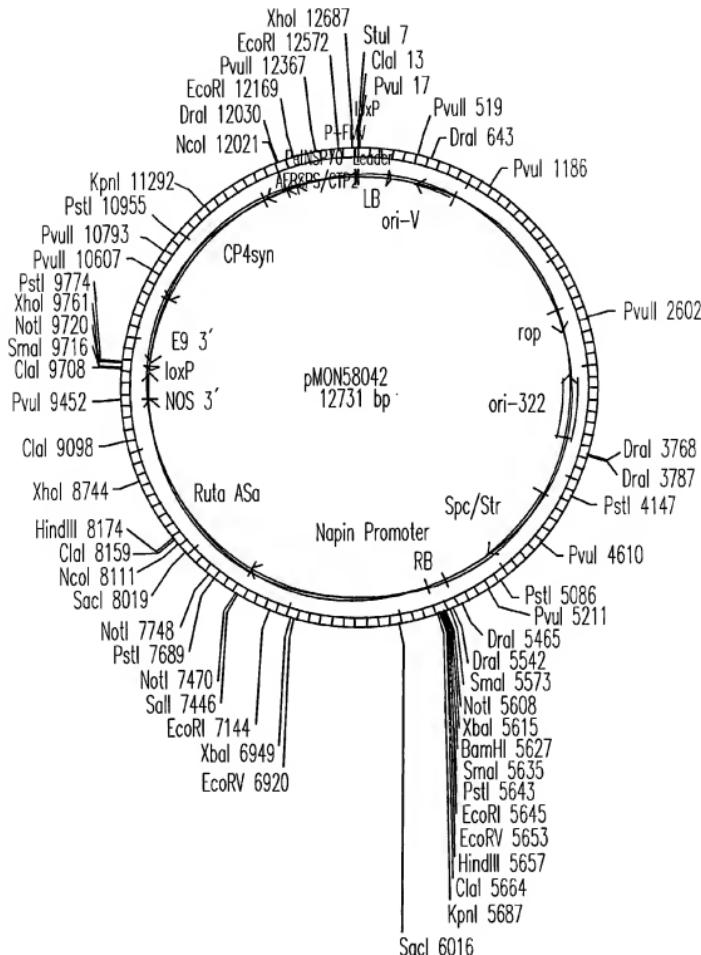


Fig. 18

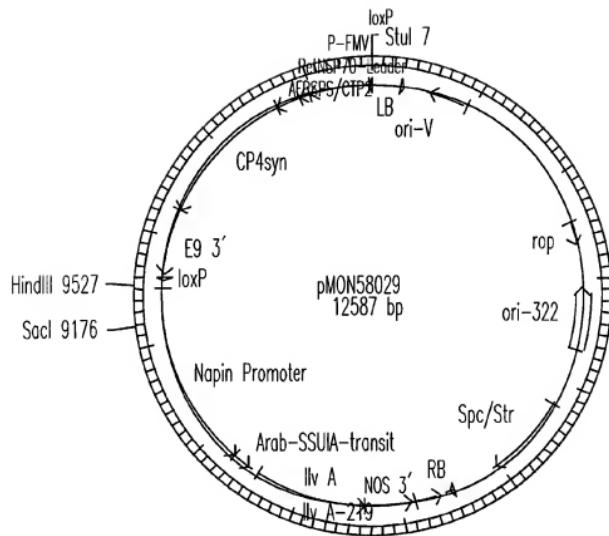


Fig. 19

21/53

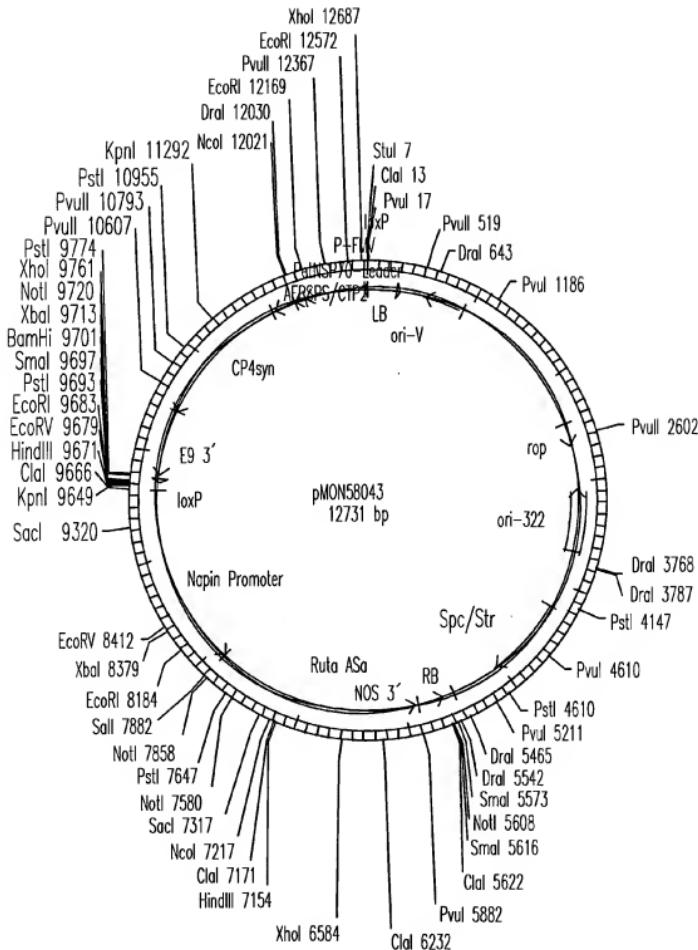


Fig.20

22/53

Fig. 21A

23/53

Fig. 21B

24 / 53

Fig. 21C

25/53

TGATVSTVRSP--VAADVFDRQFDLVLSPGPSPT--DEFDCATIKAA 596
TGASVTIVRTP--VAEEIFDRVKEDLVLVSPGPOTPK--DEFDCATIKKA 596
LGSPYIVTRNDEISIKGERIDPDRLLIISPGPCTPERKEDLGVSLDVVKY 73
LGCHFEVYRNLDEEELKKNNPRGVLSPGTPQ--DSGSIQATVLE 141

TRPREG AGRTU MONSANTO
TRPREG RHIME A30904
TRPREG SULSO B40635
TRPREG ARATH AAA32742

TIRPREG AGRTU MONSANTO
TIRPREG RHIME A30904
TIRPG SULSO B40635
TIRPG ARATH AAA32742

TERREG_AGRU_MONSANTO
TERREG_RHIME_A30904
TERREG_EULUS_B40635
TERREG_ARATH_AAA32742

Fig. 21D

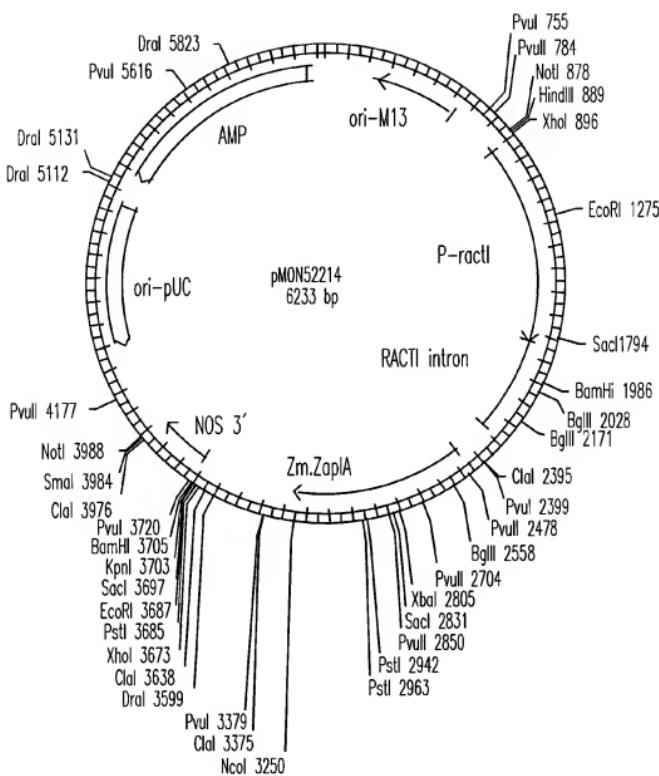


Fig. 22

27/53

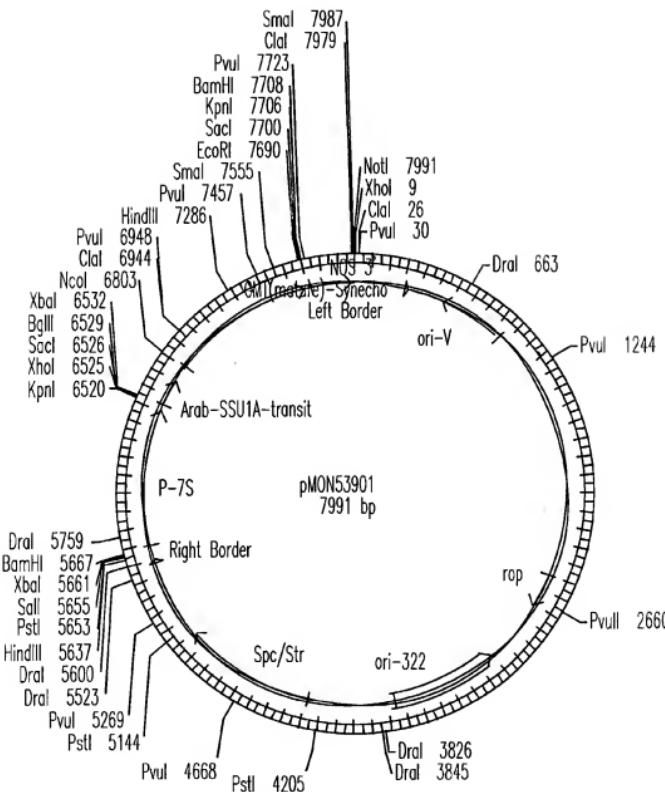


Fig.23

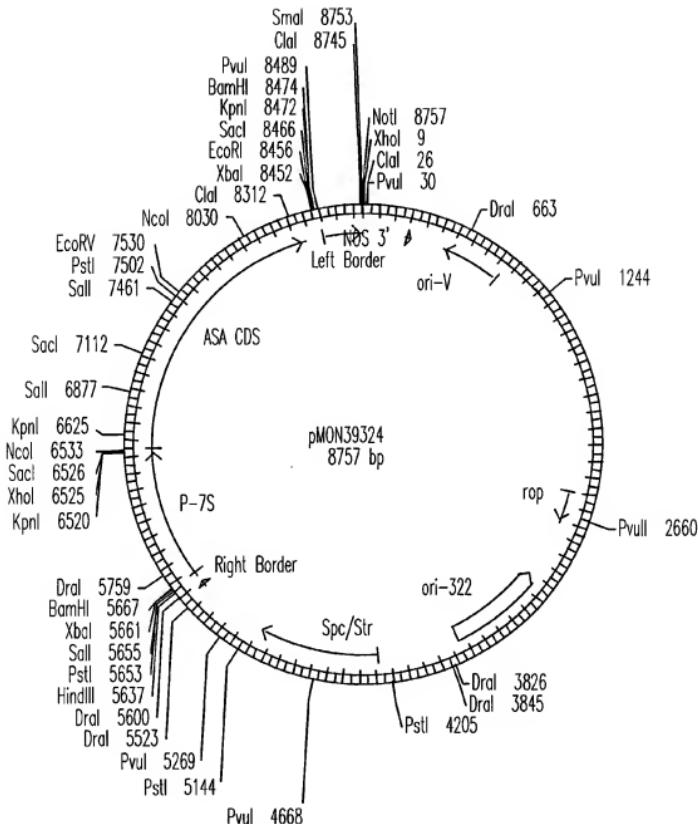


Fig. 24

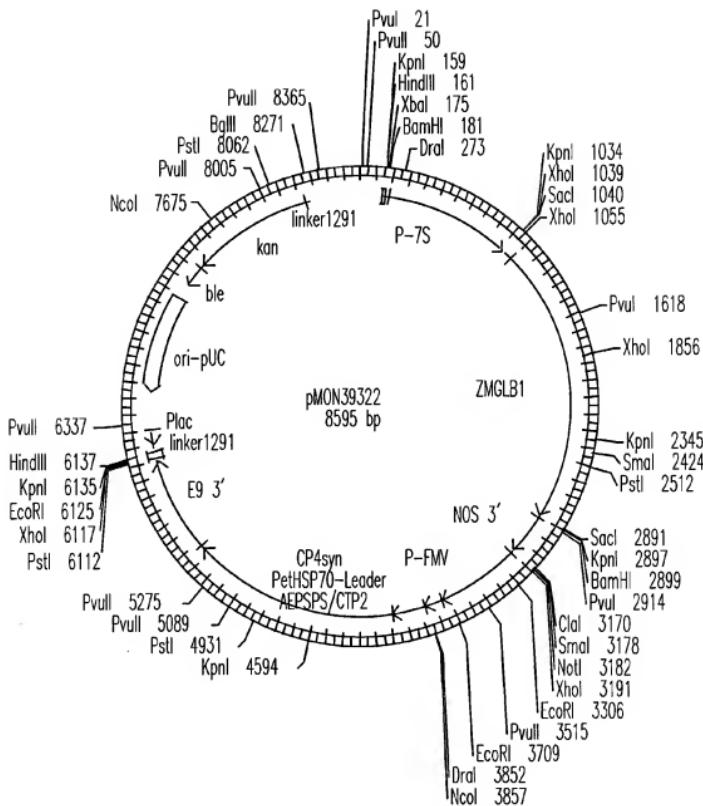


Fig. 25

30/53

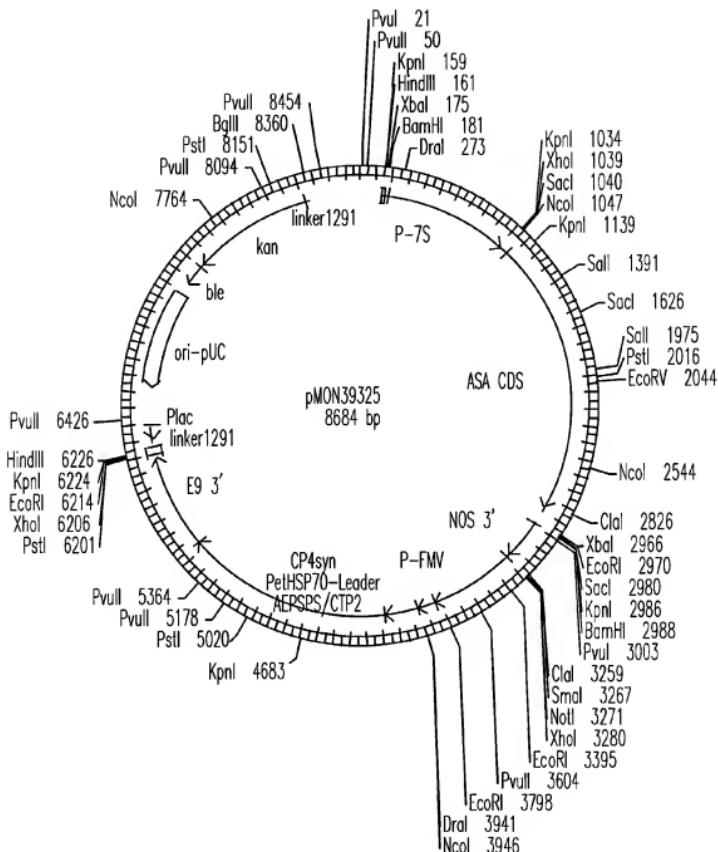


Fig.26

31/53

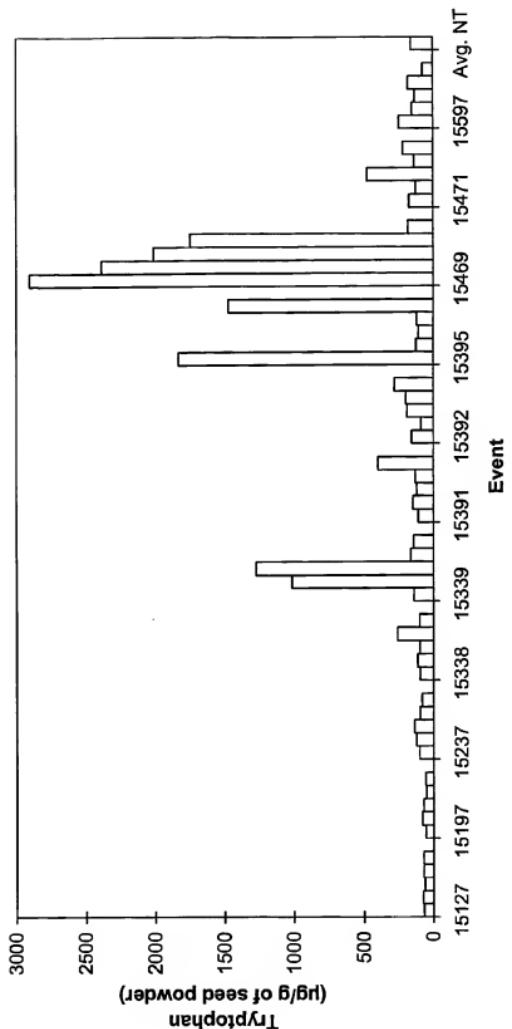


Fig. 27

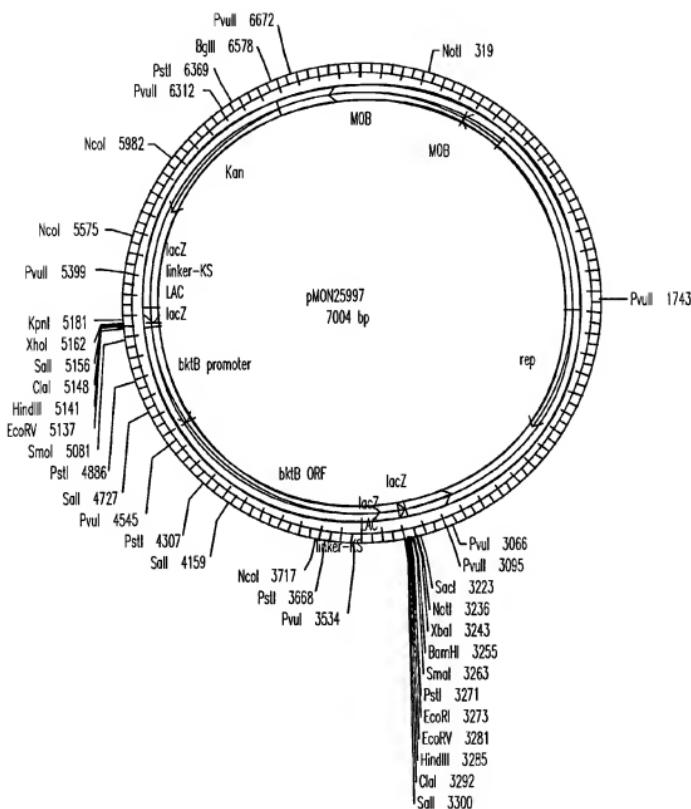


Fig.28

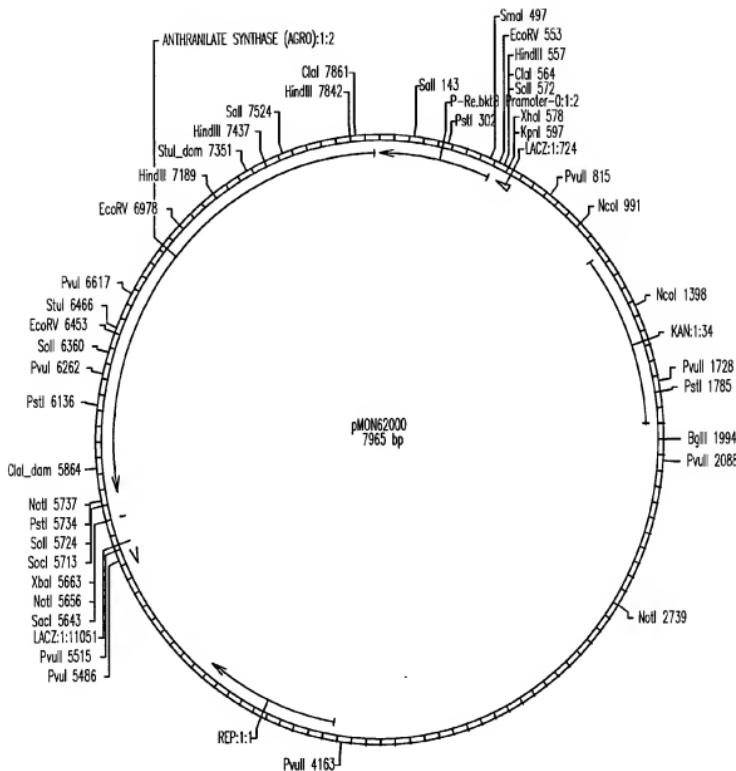


Fig.29

atgcaaacacaaaaccgactctcgaaactcggaattcctgggtggaaaacgggtatcgccaccgt
gcaaggcggtgtcggttagtcctgttccgcagtgcggaaagccgacgaaaccgcgt
acaaaggcccgcgtgtactgcgcgttatggccaccgcgtcatgcacaggagactttctga
tggctggacattctgtgtcgataatatcgactctttacgtacaacccgttgccaggatcgat
gcgca

34/53

Fig. 30

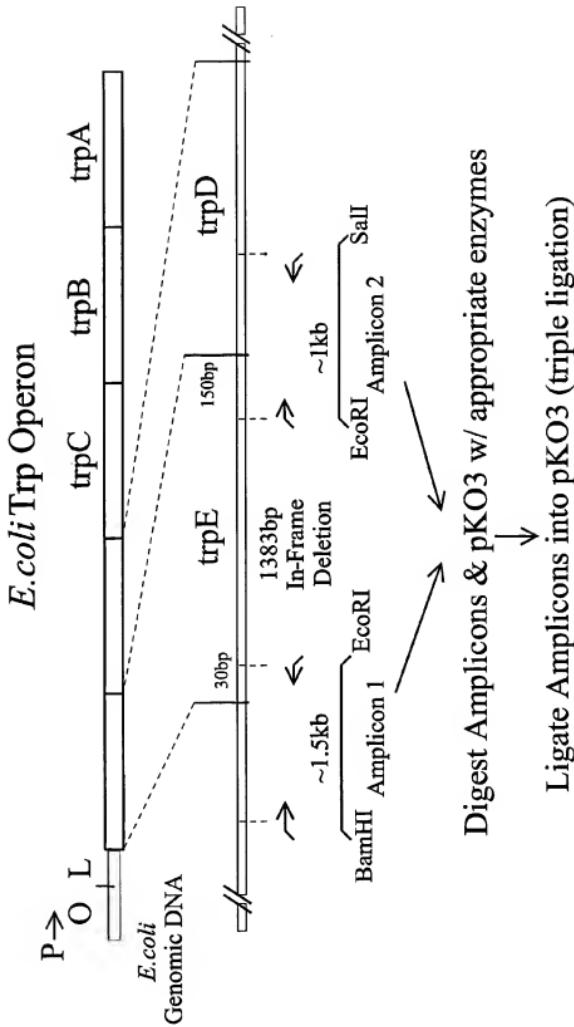


Fig. 31

36/53

- 1 - ATGGTAACGGATCATTCAAGGTGACGGGGAGACCTACGGAGACGAAAGGGGGCATCCAG - 60
 2 - M V T I Q D D G A E T Y E T K G G I Q - 61
 3 - GTCCAGCCGAAGGGCGCCACCGGATTATGCCAACGCCATTCATCGAAAG - 120
 4 - V S R K R R P T D Y A N A I D N Y I E K - 121
 5 - CTTGATTCCCATCGGGCGGGTTTTCGTCAACTATGGAAATAATCGGGCGTTACACC - 180
 6 - L D S H R G A V F S S N Y E P G R Y T - 181
 7 - CGCTGGGATACGGCCATCGTCGATCCGCCGCTGGCATTCTGTTTGGCCGCAGATG - 240
 8 - R W D T A I V D P P L G I S C F G R K M - 241
 9 - TGGATCGAAGGCCATAATGGCCGGCGGAAGTGTGCTCGATTTCATCGGAAAAGCTG - 300
 10 - W I E A Y N G R G E V L L D F I T E K L - 301
 11 - AAGGGACACCCGATCTACCCCTCGGGCTTCTCGACCCGGGCTGATCTACCGTC - 360
 12 - K A T P D L T L G A S S T R R L D L T V - 361
 13 - AACGAACGGGACCGGTCTTACCGAAGAAACGGTCTGAAAAATCCGACGGTCTCACCC - 420
 14 - N E P D R V F T E E R S K I P T V F T - 421
 15 - GCTCTCAGGCCCATCGTCGACCTCTCTATTGAGGCCGGATTGGCCATCGGGCTGTTC - 480
 16 - A L R A I V D L F Y S S A D S A I G L F - 481
 17 - GGTGCTTGGGTACGATCTGCCCTCAGTGGCTGACGGATCAAGCTTCTGGGGCGT - 540
 18 - G A F G Y D L A F Q F D A I K L S L A R - 541
 19 - CGGAAGACCGGGTGACATGGTGTGTTCTGCCCGATGAAATCCTGTGTTGATCAC - 600
 20 - P E D Q R D M V L F L P D E I L V V D H - 601
 21 - TATTCCGGCCAAGGCCCTGGATCGACGGTTACGATTTCGAGAAGGGACGGCATGACGAGGAC - 660
 22 - Y S A K A W I D R Y D F E K D G M T T D - 661
 23 - GGCAAATCCCTCCGACATTACCCCCGATCCCTCAAGGACCAACCGATAACCATCCGCCAAG - 720
 24 - G K S S D I T P D P F K T D T I P P K - 721
 25 - GGCAGATCACCCTCCGGCGAATATCCGAGGTGTTGGTAAGGCCAAGGAAAGCTCCGC - 780
 26 - G D H R P G E Y S E L V V K A K E S F R - 781
 27 - CGGGCGACCTGTTGAGGTGCTTCCGGCCAGAAATTCAATGGAGGTTGGCAAAGCAAT - 840
 28 - R G D L F E V V P G Q K F M E R C E S N

Fig. 3A

37/53

841 - CCGTCGGGATTCCGCCCTGAAGGGATCAACCCGTGCCCTATTCCTTCTTCATC - 900
 901 - P S A I S R R L K A I N P S P Y S F F I
 961 - AAATCTGGCGATCAGGAATATCTGGTGGCGCTCTGGCGAAATGTTCTGCGCGCTCTCC - 960
 961 - N L G D Q E Y L V G A S P E M F V R V S
 961 - GGCGGTGGCATGGACCTGGCCGGATATCAGGCACCATAAGCGGGGACGATCCGATT - 1020
 1021 - G R R I E T C P I S G T I K R G D D P I
 1021 - GCCGACAGCGAGGAGATTTGAAACTGTCATACTGAAAAGGAAACTGACGCC - 1080
 1081 - A D S E Q I L K L N S K K D E S E L T
 1081 - ATGGGCTCGGACCTGGACCGCAACGACAAGGCCGCTGAGCGGGTTGGTGAAG - 1140
 1141 - M C S D V D R N D K S R V C E P G S V K
 1141 - GTCAATTGGCCGGCGCAGATCGAGATGTATTCAACGCCCTCATCCACACCGTCGATCACATC - 1200
 1201 - V I G R R Q I T E M Y S R L I H T V D H I
 1201 - GAAGGGCCCTGGCGGAGGATATGGACGGCTTCTGACGGTTCTGAGCGCTGGGCC - 1260
 1261 - E G R L R D D M D A F D G F L S H A W A
 1261 - GTCACCGPTCACCGGTGACCAAAGCTGGGGCATGCGCTCATGAAAGGTATGAAAG - 1320
 1321 - V T V T G A P K L W A M R F I E G H E K
 1321 - AGCCCCGGCCCTGGATGGCGGATGGCTGGCATGGCTGGGACATGGACATGAAT - 1380
 1381 - S P R A W Y G G A I G M V G F N G D M N
 1381 - ACCGGGCTTGACGGCACCATTCGGATCAAGGGACGGTATTGGCGGAAGTGGCGGGC - 1440
 1441 - T G L T L R T I R I K D G I A E V R A G
 1441 - GCGACCCCTGCTCAATGATTCACCCGGAGGAAGAAGCCGAAACCGAACGTGAGGCC - 1500
 1501 - A T L L N D S N P Q E E A E T E L K A
 1501 - TCCGCCATGATATCAGCCATTCTGGTACGGCAAAAGGCACCAACTCTGCGCCACCAAGCGGT - 1560
 1561 - S A M I S A I R D A K G T N S A A T K R
 1561 - GATGCCGCCAAAGTGGCACCGGGTCAAGATCTGCTGACCAAGAACGACTTC - 1620
 1621 - D A A K V G T G V K I L L V D H E D S F
 1621 - GTGACACGCTGGCAATTATTCCGCCAGACGGGCGACGGTCTGACCGTCAAGATCA - 1680
 - V H T L A N Y F R Q T G A T V S T V R S

Pig. 2B

38/53

- 1681 - CCGGTCGGCAGCCGACGTGTTGATCGCTTCCAGCCGGACCTCGTTGTCGCCCCGGA - 1740
 - P V A A D V F D R F Q P D L V V I S P G
- 1741 - CCCGGCAGCCGGACGGATTTCGACTGCAAGGCAAACGATCAAGGCCGCGCCCGCGAT - 1800
 - P G S P T D F D C K A T I K A A R A R D
- 1801 - CTGCCGCAATTCGGCGTTGCCTCGGCTGCAGGCATTGGCAGAAAGCCATATGGCGGGAG - 1860
 - L P I F V C L G L Q A A L A E A Y G G E
- 1861 - CTGCCGCAAGCTTGCTGGCATCGACGCCATGCCGCTTCGGCATCCGGTGTGGAAACCC - 1920
 - L R Q L A V P M H G K P S R I R V L E P
- 1921 - GGCTCTGTTCTCCGGTCTCGGAAGGTCACGGTCGTTACCATTCGATCTTC - 1980
 - G L V F S G L G K E V T V G R Y H S I F
- 1981 - GCCGATCCGCCACCCCTGCCGCTGTTGATTTCATCATCACCGCAGAAAGCAGGGACGGACG - 2040
 - A D P A T L P D F I I T A E S E D G T
- 2041 - ATCATGGCATCGAACACGCCAAGGAACCGGTTGGCCGGTTCAAGTTCCACCCGGATTCG - 2100
 - I M G I E H A K E P V A A V Q F H P E S
- 2101 - ATCATGACGGTCCGGACAGGACGGCATGGGATGATCGAGAATGTCGTGGCATCTG - 2160
 - I M T L G Q D A G M R M I E N V V H L
- 2161 - ACCCGCAAGGCAAGAACCAAGGCCGTGA - 2190
 - T R K A K T K A A *

Hij. 32C

1	ATGGAATCCC	TAGCCGCCAC	CTCCGTGTTC	GCGCCCTCCC	GCGTGCCTGT
51	CCCGGGCGCG	CGGGCCCTGG	TTAGGGCGGG	GACGGTGTA	CCAACCAGGC
101	GGACGAGCAG	CCGGAGCGGA	ACCAGGGGG	TGAAATGCTC	TGCTGCCGTG
151	ACGGCCGAGG	CGAGGCCAGT	GATTAGCAGG	AGCGCTGGGG	CGGGGAAGGC
201	GGCGGAGGAG	GACAAGAGGC	GGTTCTTCGA	GGCGGGGGCG	CGGGGGAGCG
251	GGAAAGGGAA	CCTGGTGCCTC	ATGTGGGAGT	GCATCGTGTGTC	GGACCATCTC
301	ACCCCCGTGTC	TGCGCTTACCG	CTGCGCTCGTC	CCCGAGGACA	ACGTGCGACGC
351	CCCCCAGCTTC	CTCTTTCGAGT	CCGGTGCAGCA	GGGGCCCCAG	GGCACCCACCA
401	ACGTCGGCCG	CTATAGCATG	GTGGGAGCCC	ACCCAGTGTAT	GGAGATTGTG
451	GCCAAAGACC	ACAAGGTTAC	GATCATGGAC	CACGAGAAGA	GCCAAGTGAC
501	AGAGCAGGTA	GTGGACGACC	CGATGCAGAT	CCCGAGGACC	ATGATGGAGG
551	GATGGCACCC	ACAGGAGATC	GACGAGCTCC	CTGAATCTT	CTCGGTGGA
601	TGGGTGGGT	TCTTTTCCCTA	TGATACGGTT	AGGTATGTTG	AGAAGAAAGAA

Fig. 32A

651	GCTACCGTTC	TCCAGTGCTC	CTCAGGACGA	TAGAACCTT	CCTGATGTGC
701	ACTTGGACT	CTATGATGAT	GTTCATGTTCT	TCGATAATGT	TGAGAAAGAA
751	GTATATGTTA	TCCATTGGGT	CAATGTGGAC	CGGCATGCCAT	CTGTTGAGGA
801	AGCATACCAA	GATGGCAGGT	CCCGACTAAA	CATGTTGCTA	TCTAAAGTGC
851	ACAATTCCAA	TGTCCCCACA	CTCTCTCCCTG	GATTGTGAA	GCTGCACACA
901	CGCAAGTTG	GTACACCTT	GAACAAAGTCG	ACCATGACAA	GTGATGAGTA
951	TAAGAATGCT	GTTCCTGCAGG	CTAAAGAACAA	TATTATGGCT	GGGGATATCT
1001	TCCAGATTGT	TTAAAGCCAG	AGGTTCGAGA	GACGAACATA	TGCCAACCCA
1051	TTTAGGTTT	ATCGAGCATT	ACGGATTGTG	AATCCTAGCC	CATACATGGC
1101	GTATGTACAG	GCAAGAGGCT	GTGTATTGGT	TGCGTCTAGT	CCTGAAATTG
1151	TTACACGAGT	CAGTAAGGGG	AAGATTATTA	ATCGACCACT	TGCTGGAACT
1201	GTTCGAAAGGG	GCAAGACAGA	GAAGGAAGAT	CAAATGCAAG	AGCAGCAACT
1251	GTAAAGTGT	AAAAAACAGT	GTGCGGAGCA	CATAATGCTT	GTGGACTTGG

Fig. 3B

1301 GAAGGAATGA TGTGGCAAG GTATCCAAAC CAGGATCAGT GAAGGGGGAG
1351 AAGTTGATGA ACATTGAGAG ATACTCCCAT GTTATGCACA TCAGCTCAAC
1401 GGTTAGTGGA CAGTTGGATG ATCATCTCCA GAGTTGGAT GCCTTGAGAG
1451 CTGCCCTTGCC CGTGGAACCA GTCAGTGGTG CACCAAAGGT GAAGGCCATG
1501 GAGTTGATTG ATAAGTTGGA AGTTACGAGG CGAGGACCAT ATAAGTTGG
1551 TCTAGGAGGA ATATCGTTG ATGGTGACAT GCAAATTGCA CTTTCTCTCC
1601 GCACCATCGT ATTCTCAACA GCGCCGAGCC ACAACACGAT GTACTCATAC
1651 AAAAGACGCAG ATAGGGCTCG GGAGGGGTC GCTCATCTTC AGGCTGGTGC
1701 AGGCATTGTT GCCGACAGTA GCCCAGATGA CGAACAAACGT GAATGCGAGA
1751 ATAAGGCTGC TGCACTAGCT CGGGCCATCG ATCTTGAGA GTCAAGCTTT
1801 GTAGACAAAG AATAG

Fig. 3C

1 MESLAATSVF APSRVAVPA A RALVRAGTVV PTRRTSSRSRSG TSGYRKCSAAV
51 TPQASPVISR SAAAAKAAEE DKRRFFEEAAA RGSGKGNLVP MWECIVSDHL
101 TPVLAYRCLV PEDNVDAEPSF LFESVEQGPQ GTTNVGRYSM VGAHPVMEIV
151 AKDHKVTIMD HEKSQVTEQV VDDPMQIPRT MMEGWHPQQI DELPESFSGG
201 WVGFESYDTV RYVEKKKLPPF SSAPQDDRNL PDVHLLGYDD VLVFDNVEKK
251 VYVIHWVNVD RHASVEEAYQ DGRSRLNMLL SKVHNNSNVPT LSPGFVKLHT
301 RKFGTPLNKS TMTSDEYKNA VLQAKEHIMA GDIFQIVLSQL RFERRTYANP
351 FEVYRALRIV NPSPYMAYVQ ARGCVLVASS PEILTRVSKG KIIINRPLAGT
401 VRRGKTEKED QMQEQQLLSD ERQCAEHIML VDLGRNDVGRK VSKPGSVKVE
451 KLMNIERYSH VMHISSTVSG QLDLHLQSWD ALRAALPVGT VSGAPKVKAM
501 ELIDKLEVTR RGPySGGLGG ISFDGDMQIA LSLRTIVFST APSHNTMSY
551 KDADRRREWV AHLQAGAGIV ADSSPDDEQR ECENKAALA RAIDLAESAFAF
601 VDKE*

Fig. 32/

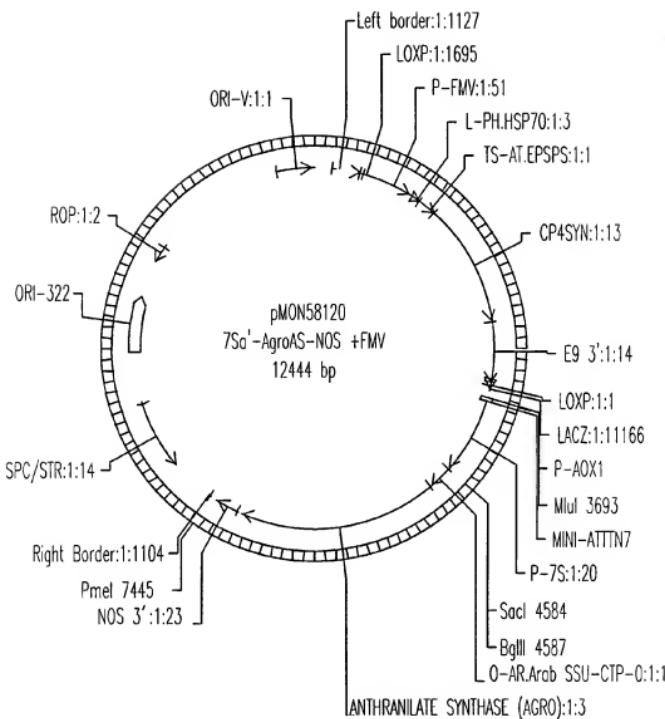


Fig.34

Agarthu	15889565	PGRYTRWDTA1VDPPGLISCFGRKWWIEAYNGRGEVILDFITEKLKATPDLTGASSSTR
Rhime	_136328	PGRYTRWDTAVVDPPPLAISSFGRSWIWEAYNERGEVILLAIDLKSVADITLGSIAARR
Meslo	_1374468	PGRYTRWDTA1DPPVLISARGRAMTLAUNRGEALLPVIGTKTGLGLADITIATTPTSL
AzBr	_1717765	PGRYRRHLAGFTDPVVALTARGTRTLRIDALNRGQVLLPAVAEALGLEAGHEAPSR
BruMe	_17986732	PGRYTRWDTA1VDPPVTTSRARTMRTEALNARGVILLREPLTDVKALSEVKIDOSGENR
Nostoc	_1722910	PGRYKRWAIAGFNPPVELSTSGNTFLTALNENERGYVLLPVIFECLSSEQLOKLTEHHHK
Nostoc	_17230725	PGRYKRWAIAGFNPPPLQTRENAAITSSLNPRGOVLLPTFOHLSAQSOLOOLSIUNHDY
RhCPa_TriPEG		PGRYFSDLGPDPPLTRAETKIEALNPRGRVLAFSLDKLEPCVVVEQCATKI

Fig. 25A

Fig. 35B

Fig. 35C

Fig. 35D

AgrTu	15889565	DAGMRMIENVVVHLTRKAKTKAA---	729
RhiMe	_136328	DAGMRMIENVVVAHLKRAKTKAA---	729
MesLo	_13472468	NAGMRIIENIVAHLPRAKEKAA---	733
AzoBr	_1717765	GAGLALLGNVMMDRLAAGALTDAAA--	732
BruMe	_17986732	NAGMRMIENIVTHLAGHKKARRTNY-	731
Nostoc	_17227910	EVGLMMIKNIVVQKYTOSOOSTVPIVD	735
Nostoc	_17230725	EVGQTIIKNVVQTYTOTLETSIYS--	715
RhoPa	<u>T</u> rPEG	EVGLRIVENAFRLGQAA-----	719
	*	:	*
			.

Hg. 32'

1 ATGGTGTACCA TCATTCAAGGA TGACGGTGGCC GAGACCTACG AGACCAAAGGG CGGCATCCAG
 6.1 GTGAGCGCA AGGGCGGCC CACCGATTAC GCGAAAGCCA TCGATAACTA CATGAAAAG
 121 CTTGATTCCTC ATTCGGGGTGC CGTGTCTTCG TCCAACCTACG AATACCCAG CGCTACACC
 181 CGCTGGATA CGGCCATCTGT CGATCCACCA CTCGGCAT CTCGGCTTCGG CGCAAGATG
 241 TGGATGAAAG CCTAACAGG CGCGGGCAA GTGCTGCTCG ATTACATTAC CGAAAAGCTG
 301 AAGGCACAC CGGATCTCAC CCTCGGGCT TCCCTACCCG GCGCCCTCGA TCTTACCGTC
 361 AACGAACAG AC CGGTCTCTT CACCGAAGAA GAACCTCCA AAATCCAC CGCTTCAC
 421 GCTCTAGGG CCATCGTCCA CCTCTCTCTAC TCCAGGCCG ATTCGGCCAT CGGCCCTGTT
 481 GGTGCTCTCG GTTACGATCT CGCCCTTCCAG TTTCGAGGCCA TCAAGCTTTC CCGGCCCCGC
 541 CCAGAGACC AGGCCCTGGAT GGTCGTTCTC CTGCCGAT AAATCTCGT GTTGTATC
 601 TACTCGCCA AGGCCCTGGAT CGACCGCTAC GATTTGAGA AGGACGGCAT GACCCACCGAC
 661 GGCAATCTT CGACATTAC CCCGGATCCC TCAAGACCA CGGATACCA CGCACCAAG
 721 GGGGATCACC GCCCCGGCGA ATACTCCGAG CTTGTGTTGA AGGGCAAGGA AGGCTTCCGC
 781 CGGGGGGACC TTGTTCGAGGT CGTTCCCGGC CAGAACTTCA TGGAGCGCTG CGAAAGCAAC
 841 CCATCGCCA TTGCGCCGAT CCTGAAGGG ATCAACCCAT CCCCTCTACTC TTCTCTCATC
 901 AACCTGGCG ATCAGGAAATA CCGGGTGGC GCCTCCCCAG AAATGTTCGT GCGGCTCTCC
 961 GGGCCGCA TGAGACCTG CCAAACTCA GGCAACATCA AGCGGGCA CGATCCAATT
 1021 GCGGACAGCG AGCAAGATTTC GAAACTGCTC AACTCCAAA AGGAGGAATC GAAACTGACC
 1081 ATGGTCTCCG AGCTGGACCG CAACGACAAG AGCGGGCTC GCGAGCCAGG TTCCGGTGAAG
 1141 GTCATGGCC GCGCCGAGT CGAGATGTC TCA CGCTCTCA TCCACACCGT CGATCACATC
 1201 GAAGGCCGCC TGCGCGACCA TATGGACGCC TTGGACGGTT TCCCTAGCCA CGCTTGGGC
 1261 GTCACTGCA CGGGTGACCC AAAGCTGTTG GCGATGGCT TCATGAAAGG TCATGAAAAG
 1321 AGCCCAAGCG CCTGGTACCG CGGTGCCATC GGCATGGCTC GCTTCACCG CGACATGAAAC
 1381 ACCGGCCCTGA CCCTGCGCAC CATCCGCATC AAGGACCGT TTGGCCGAAGT GCGGCGCCGGC
 1441 GCCAACCTGC TCAACGATTC CAACCCACAG GAAGAGGAAG CGAAAAACCGA ACTGAAAGGCC
 1501 TCCCGCATGA TCTCAGGCCA TCGGACGCA AAMGGACCA ACTCTGCGGC CACCAAGGCC
 1561 GATGCCGCCA AAGTGGGCCA CGGGGTCAAG ATCCCTGCTCG TCGACCCAGA AGACAGCTTC

H. Jia

1621 GTGACACACC TGGCCAACTA CTTCCGCCAG ACCGGGCCA CCGGTCTCCA CGTCAGGTCA
 1681 CCAGTGGAG CGAGCGTGT CGATCGCTT CGACCGATT CAACCGATT CAACCGATT CGACTGCAAG GCACCATCA AGGGCGCCCG CGCCCGCGAT
 1741 CCCGGAGCC TCGGCATCT CGACTGGCTT CCTGGCTTG CAGGCATTGG CAGGAAGCTTA CGAGGCTTA CGGGCGGAG
 1801 CTGCCAATCT TCGGCATCT TTGGCTGTGCC CATGCAAGGC AAGCCTTCCC GCATCCCGCTT GCTGGAACCC
 1861 CTGCCCCAGC TTGGCTGTGCC GGCAAGGAA GTCAAGGAA GTCAAGGAA GTCAAGGAA GTCAAGGAA GTCAAGGAA
 1921 GGCCCTGTCT TCTCCGGTCT CGACCGATT CCACCCCTGCC ACGGGATTIC ATCATACCC CAGAAAGCGA GAACGGCAC
 1981 GCCCATCCCG CAAGGAACCA GTGGCCGCCG TTCAAGTTCCA CCCAGAACCC
 2041 ATCATGGCA TCGAACACGCC CAAGGAACCA GTGGCCGCCG TTCAAGTTCCA CCCAGAACCC
 2101 ATCATGACCC TCGGTCAAGGA CGCCGGCATG CGCATGATCG AGAACGTCTG GTGCACTTG
 2161 ACCGGCAAGG CCAAGACCAA GGCCGCCCTGA

Fig. 3B

Fig. 32A

Fig. 32B

1990	12000	2010	2020	2030	2040	2050	2060	2070	2080	2090
GC	CG	AT	CC	CT	GG	TC	GA	AC	GG	CC
GC	GT	AT	CC	CT	GG	TC	GA	AC	GG	CC
GC	GC	AT	CC	CT	GG	TC	GA	AC	GG	CC

Fig. 32C

SEQUENCE LISTING

<110> Monsanto
5 Renessen LLC
 Weaver, L.M.
 Liang, J.
 Chen, R.
 Jeong, S.S.
10 Mitsky, T.
 Slater, S.
 Rapp, W.

<120> Transgenic High Tryptophan Plants
15
<130> 1463.002WO1

<150> US 60/288,904
<151> 2001-05-04
20
<160> 103

<170> FastSEQ for Windows Version 4.0

25 <210> 1
<211> 2190
<212> DNA
<213> Agrobacterium tumefaciens

30 <400> 1
atggtaacga tcattcagga tgacggagcg gagacctaeg agacgaaagg cggcatccag 60
gtcagccgaa agcgcggccc cacggattat gcacgcgcgc tcgataatta catcgaaaag 120
cttggatccc atcgccgcgc ggtttttcgc tccaactatg aataccggg cggttacacc 180
cgctggata cggccatcgatcgatccgcgc ctccggatcc ttgttttgg ccgcaagatg 240
35 tggatcgaaatgg cctataatgg cgcggcgaa gtgtcgatcg atttcattac gaaaaagctg 300
aaggcgacac ccgtatctac cctccggcgatcgatccgcgc ttttacatcgatcgatcg 360
aacgaaccgg acggatgttccat caccgaagaa gaacgtcgaa aaatccggac ggttttcc 420
gctctcagatcgatcgatccgcgc ttgttttgcgc ttcggatccat ccggatcgatcgatcg 480
40 ggtgccttcgc gttacgtatcgatccgcgc ttgcacgcgc tcaagatccatcgatcgatcg 540
ccggaaagacc agcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg 600

2

tattccggcca	aggcctggat	cgaccgttac	gatttcgaga	aggacggcat	gacgacggac	660	
ggcaaatactt	ccgcacattac	ccccgatccc	ttcaagacca	ccgataccat	ccgcggccaag	720	
ggcgatcacc	gtccccggca	atattccag	cttgggttga	aggccaagga	aagcttccgc	780	
5	cgcggcacc	tgttccgggc	cagaatacc	tggagcgttg	cgaaaagcaat	840	
cgcggcacc	tttcccgccg	cctgaaggcg	atcaaccgt	cgccttattc	cttcttcattc	900	
aactctccggc	atcaggaata	tctggtcggc	gcctcgccgg	aaatgttctgt	gcccgttcc	960	
ggcgcgtcgca	tcgagacccgt	cccgatatac	ggcacccatca	agecgccgca	cgatccgatt	1020	
gcgcacagcg	aggagattt	gaaatgttc	aactcgaaaa	aggacgatcc	cgaacttgacc	1080	
atgtgtccgg	acgttggaccg	caacgacaag	aggccgtct	gcgagccggg	ttcgggtgaag	1140	
10	gtcattggcc	gcgcggcagat	cgagatgtat	tcacgcgttca	tcacacccgt	cgatcacatc	1200
gaaggccggc	tgcgcgacga	tatggaccc	tttgcgggtt	tcctcagccca	ccgcctggggcc	1260	
gtacccgtca	ccgggtcacc	aaagtgtgg	gocatgggt	tcatcgaaagg	tatgaaaag	1320	
agccccggcg	cctggatattg	cggttgcgatc	ggcatggtc	gcttcacacgg	cgacatgaat	1380	
acccggctca	cgctcgccac	catccggatc	aaggccgtta	ttgcgcgaaat	gcccgcggcc	1440	
15	gogacccgtc	tcaatgttcc	caaccgcac	gaagaagaag	ccgaaaccgt	actgaaggcc	1500
tccgcatcg	tatcagccat	tcgtgacgc	aaaggccacca	actctgcgc	caccaagcgt	1560	
gatgccgc	aagtccggc	cggcgctcaag	atccctgttc	tcgaccacga	agacagcttc	1620	
gtgcacacgc	tggcgaatta	tttccgcgc	acggccgcga	cggctctgc	cgtcagatca	1680	
ccggcgcgc	ccgcacgttt	cgatcgcttc	cagccggacc	tcgttgttct	gtcgcggcga	1740	
20	ccccgcagcc	cgacggattt	cgactgcaag	gcaacgtatca	aggccgcggcg	cgcccgccgat	1800
ctgcgcgatet	tctgggtttt	cctcggtctg	caggcatgg	cagaaggctca	tggccggcgg	1860	
ctgcgcgc	ttgtgtgtcc	catgcacgc	aaggcttcgc	gcatccgcgt	gtggaaaccc	1920	
ggccctgtt	tctccggctt	cggaaggaa	gtcacggtc	gtcggttacca	ttcgatcttc	1980	
gcccgcgc	ccacccctgc	gctgttattc	atcatccac	cagaaagcga	ggacggcacc	2040	
25	atcatggca	tgcacacgc	caaggaaacc	gtggccgcgc	ttagtgcata	ccccgaatcg	2100
atcatgcgc	tgcacacgc	cgccggatcg	cggtatcg	agaatgttgc	gtgtgcatctg	2160	
acccgc	cgaaaggccaa	ggccgcgtga				2190	

<210> 2

30 <211> 1815

<212> DNA

<213> Zea mays

<400> 2

35	atggaaatccc	tagccgcac	ctccgtgttc	gcgcctccc	gctgcgcgcgt	ccoggccggcg	60
cgggccctgg	ttagggccgg	gacgggttca	ccaaaccaggc	ggacggacgc	ccggacggca	120	
accagcgggg	tgaaatgttc	tcgtgcgttgc	acgcgcgcgg	cgagcccaat	gattacggagg	180	
acgcgtcgcc	cggcgtcgcc	ggccggaggag	gacaaggaggc	ggttcttcga	ggccggccggcg	240	
cggggggaggc	ggaaggggaa	cctgggttcc	atgtggggat	gcatcgatc	ggaccatctc	300	
40	accccccgtgc	tgcgttccat	ctgcgttccat	cccgaggaca	acgtcgacgc	ccccagatcc	360

3

ctcttcgagt	ccgtcgagca	ggggccccag	ggcaccacca	acgtcgcccg	ctatagcatg	420	
gtgggagccc	accagtgtat	ggagattgtg	gccaaaagacc	acaagggtac	gatcatggac	480	
cacgagaaga	gccaagtgcac	agagcaggtt	gtggacgacc	cgatgcagat	cccgaggacc	540	
atgatgggg	gatggccccc	acagcagatc	gacgagttcc	ctgaatctt	ctccggtgaa	600	
5	tgggtgggtt	tctttccata	tgatacggtt	aggtatgtt	agaagaagaa	gctaccgttc	660
tccagtgcctc	tcaggacgca	taggaaacctt	cctgatgtgc	acttgggact	ctatgtatgt	720	
gttcttagtct	tcgataatgt	tgagaagaaa	gratatgtta	tccattgggt	caatgtggac	780	
cggcatgcat	ctgttgggaa	agcataccaa	gatggcagggt	cccgactaaa	catgttgcta	840	
tctaaatgtc	acatccaa	tgtccccca	ctctcttcgt	gatttggaa	gctgcacaca	900	
10	cgaagggtt	gtacacccccc	gaacaagtgc	accatgacaa	gtgatgagta	taagaatgt	960
gtttcgcgg	ctaaggaaaca	tattatgggt	ggggatatct	tccagatgtt	ttaaggccag	1020	
aggttcgaga	gacgaaacata	tgcacacca	tttgaggattt	atcgagcatt	acggattgtt	1080	
aatcccttag	catacatggc	gtatgtacag	gcaagagggt	gtgttattgtt	tgctgtctagt	1140	
cctgaaatcc	ttacacgagt	cgtaaagggg	aagattatta	atcgaccact	tgctggact	1200	
15	gttgcgagggg	gcaagacaga	gaaggaaagat	caaatgcga	agcagcaact	gttaatgtat	1260
gaaaaaacatg	gtggcgacga	cataatgttt	gtggacttgg	gaaggaaatga	tgttggcaag	1320	
gtatccaaac	caggatcgt	gaagggtggag	aagttgtatga	acattggagag	atactcccat	1380	
gttatgcaca	tcagctcaac	gtttagtggaa	cagttggat	atcatcttca	gagttgggat	1440	
gccttggag	ctgccttgc	cggttggaaaca	gtcagttggt	ccacaaagggt	gaaggccatg	1500	
20	gagttgattt	ataagttggaa	agttacgagg	cgaggacat	atagtgggtt	tcttagggaga	1560
atatcgttt	atgggtgacat	gcaaatttgc	ctttcttc	gcaccatcgt	attctcaaca	1620	
gcgcgcgagcc	acaaacacgt	gtactcatac	aaagacgcag	ataggcgatcg	ggagttgggtc	1680	
gctcatcttc	aggctgggtc	aggcatgtt	ggccgacagta	gccccagatga	cgaacaaacgt	1740	
gaatgcgaga	ataaggctgc	tgcactatgt	cggggccatcg	atcttgcaga	gtcagctttt	1800	
25	gtgacacaaag	aataat				1815	

<210> 3

<211> 1993

<212> DNA

30 <213> Ruta graveolens

<400> 3							
aaaaaaaaatcg	tctgtttttc	gtgtttggac	atttcagcg	cactgggtgc	catcagttga	60	
ttcgactcat	ttgattttt	ttgttgttgc	gcataggttgc	cagcggcaac	gtcgatgca	120	
35	tcccttaaat	tcttccaaac	tctgtttccca	cccgatcg	gtctgttcc	ggttccgaaac	180
aatgtcacat	gcaataacat	ccccaaatgtt	cgacgtcccg	tccggacagt	caaatgtgc	240	
gtttttccat	gaaacagttac	catcaacggc	ggggccgcca	cgaccaacgg	tgcgatccgc	300	
gcccgttaac	gcgcacatcc	gaccacact	acatatgtt	gtgtatgcac	cagattttatc	360	
gacttcttca	aaaggccaaa	tcttagtgc	ttataccgtt	gcataattgc	ggatcatctc	420	
40	acgcgggtgc	ttgcctatag	atgtttgggtt	caagaagac	ataaaagac	tccaaatgtt	480

4

53102-4

5211-729

30-52123 PBT

<213> *Agrobacterium tumefaciens*

5600> 4

Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys

35 1 5 10 15

le Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Ty:

20 25 30

sp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg

35 40 45

40 Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr

5

50	55	60
Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met		
65	70	75
Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile		80
5	85	90
Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser		95
100	105	110
Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr		
115	120	125
10 Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala		
130	135	140
Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe		
145	150	155
Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu		160
15	165	170
Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro		175
180	185	190
Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp		
195	200	205
20 Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser		
210	215	220
Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys		
225	230	235
Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys		240
25	245	250
Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys		255
260	265	270
Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu		
275	280	285
30 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp		
290	295	300
Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser		
305	310	315
Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly		320
35	325	330
Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Asn Ser		335
340	345	350
Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn		
355	360	365
40 Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg		

370	375	380
Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile		
385	390	395
Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser		400
5	405	410
His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met		415
420	425	430
Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly		
435	440	445
10 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr		
450	455	460
Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly		
465	470	475
Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Glu Ala Glu Thr		480
15	485	490
Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly		495
500	505	510
Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly		
515	520	525
20 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu		
530	535	540
Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser		
545	550	555
Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val		560
25	565	570
Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr		575
580	585	590
Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu		
595	600	605
30 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu		
610	615	620
Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro		
625	630	635
Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr		640
35	645	650
His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile		655
660	665	670
Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys		
675	680	685
40 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu		

690	695	700	
Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu			
705	710	715	720
Thr Arg Lys Ala Lys Thr Lys Ala Ala			
5	725		
<210> 5			
<211> 604			
<212> PRT			
10 <213> Zea mays			
<400> 5			
Met Glu Ser Leu Ala Ala Thr Ser Val Phe Ala Pro Ser Arg Val Ala			
1	5	10	15
15 Val Pro Ala Ala Arg Ala Leu Val Arg Ala Gly Thr Val Val Pro Thr			
20	25	30	
Arg Arg Thr Ser Ser Arg Ser Gly Thr Ser Gly Val Lys Cys Ser Ala			
35	40	45	
Ala Val Thr Pro Gln Ala Ser Pro Val Ile Ser Arg Ser Ala Ala Ala			
20	50	55	60
Ala Lys Ala Ala Glu Glu Asp Lys Arg Arg Phe Phe Glu Ala Ala Ala			
65	70	75	80
Arg Gly Ser Gly Lys Gly Asn Leu Val Pro Met Trp Glu Cys Ile Val			
85	90	95	
25 Ser Asp His Leu Thr Pro Val Leu Ala Tyr Arg Cys Leu Val Pro Glu			
100	105	110	
Asp Asn Val Asp Ala Pro Ser Phe Leu Phe Glu Ser Val Glu Gln Gly			
115	120	125	
Pro Gln Gly Thr Thr Asn Val Gly Arg Tyr Ser Met Val Gly Ala His			
30	130	135	140
Pro Val Met Glu Ile Val Ala Lys Asp His Lys Val Thr Ile Met Asp			
145	150	155	160
His Glu Lys Ser Gln Val Thr Glu Gln Val Val Asp Asp Pro Met Gln			
165	170	175	
35 Ile Pro Arg Thr Met Met Glu Gly Trp His Pro Gln Gln Ile Asp Glu			
180	185	190	
Leu Pro Glu Ser Phe Ser Gly Gly Trp Val Gly Phe Phe Ser Tyr Asp			
195	200	205	
Thr Val Arg Tyr Val Glu Lys Lys Lys Leu Pro Phe Ser Ser Ala Pro			
40	210	215	220

Gln Asp Asp Arg Asn Leu Pro Asp Val His Leu Gly Leu Tyr Asp Asp
 225 230 235 240
 Val Leu Val Phe Asp Asn Val Glu Lys Lys Val Tyr Val Ile His Trp
 245 250 255
 5 Val Asn Val Asp Arg His Ala Ser Val Glu Glu Ala Tyr Gln Asp Gly
 260 265 270
 Arg Ser Arg Leu Asn Met Leu Leu Ser Lys Val His Asn Ser Asn Val
 275 280 285
 Pro Thr Leu Ser Pro Gly Phe Val Lys Leu His Thr Arg Lys Phe Gly
 10 290 295 300
 Thr Pro Leu Asn Lys Ser Thr Met Thr Ser Asp Glu Tyr Lys Asn Ala
 305 310 315 320
 Val Leu Gln Ala Lys Glu His Ile Met Ala Gly Asp Ile Phe Gln Ile
 325 330 335
 15 Val Leu Ser Gln Arg Phe Glu Arg Arg Thr Tyr Ala Asn Pro Phe Glu
 340 345 350
 Val Tyr Arg Ala Leu Arg Ile Val Asn Pro Ser Pro Tyr Met Ala Tyr
 355 360 365
 Val Gln Ala Arg Gly Cys Val Leu Val Ala Ser Ser Pro Glu Ile Leu
 20 370 375 380
 Thr Arg Val Ser Lys Gly Lys Ile Ile Asn Arg Pro Leu Ala Gly Thr
 385 390 395 400
 Val Arg Arg Gly Lys Thr Glu Lys Glu Asp Gln Met Gln Glu Gln
 405 410 415
 25 Leu Leu Ser Asp Glu Lys Gln Cys Ala Glu His Ile Met Leu Val Asp
 420 425 430
 Leu Gly Arg Asn Asp Val Gly Lys Val Ser Lys Pro Gly Ser Val Lys
 435 440 445
 Val Glu Lys Leu Met Asn Ile Glu Arg Tyr Ser His Val Met His Ile
 30 450 455 460
 Ser Ser Thr Val Ser Gly Gln Leu Asp Asp His Leu Gln Ser Trp Asp
 465 470 475 480
 Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly Ala Pro Lys
 485 490 495
 35 Val Lys Ala Met Glu Leu Ile Asp Lys Leu Glu Val Thr Arg Arg Gly
 500 505 510
 Pro Tyr Ser Gly Gly Leu Gly Gly Ile Ser Phe Asp Gly Asp Met Gln
 515 520 525
 Ile Ala Leu Ser Leu Arg Thr Ile Val Phe Ser Thr Ala Pro Ser His
 40 530 535 540

9

Asn Thr Met Tyr Ser Tyr Lys Asp Ala Asp Arg Arg Arg Glu Trp Val
545 550 555 560
Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ser Pro Asp
565 570 575
5 Asp Glu Gln Arg Glu Cys Glu Asn Lys Ala Ala Ala Leu Ala Arg Ala
580 585 590
Ile Asp Leu Ala Glu Ser Ala Phe Val Asp Lys Glu
595 600

10 <210> 6
<211> 613
<212> PRT
<213> Ruta graveolens

15 <400> 6
Met Ser Ala Ala Ala Thr Ser Met Gln Ser Leu Lys Phe Ser Asn Arg
1 5 10 15
Leu Val Pro Pro Ser Arg Arg Leu Ser Pro Val Pro Asn Asn Val Thr
20 25 30
20 Cys Asn Asn Leu Pro Lys Ser Ala Ala Pro Val Arg Thr Val Lys Cys
35 40 45
Cys Ala Ser Ser Trp Asn Ser Thr Ile Asn Gly Ala Ala Ala Thr Thr
50 55 60
Asn Gly Ala Ser Ala Ala Ser Asn Gly Ala Ser Thr Thr Thr Thr
25 65 70 75 80
Tyr Val Ser Asp Ala Thr Arg Phe Ile Asp Ser Ser Lys Arg Ala Asn
85 90 95
Leu Val Pro Leu Tyr Arg Cys Ile Phe Ala Asp His Leu Thr Pro Val
100 105 110
30 Leu Ala Tyr Arg Cys Leu Val Gln Glu Asp Asp Lys Glu Thr Pro Ser
115 120 125
Phe Leu Phe Glu Ser Val Glu Pro Gly Arg Ile Ser Thr Val Gly Arg
130 135 140
Tyr Ser Val Val Gly Ala His Pro Val Met Glu Val Ile Ala Lys Asp
35 145 150 155 160
Asn Met Val Thr Val Met Asp His Glu Lys Gly Ser Leu Val Glu Glu
165 170 175
Val Val Asp Asp Pro Met Glu Ile Pro Arg Arg Ile Ser Glu Asp Trp
180 185 190
40 Lys Pro Gln Ile Ile Asp Asp Leu Pro Glu Ala Phe Cys Gly Gly Trp

10

195	200	205
Val Gly Phe Phe Ser Tyr Asp Thr Val Arg Tyr Val Glu Lys Lys Lys		
210	215	220
Leu Pro Phe Ser Lys Ala Pro Gln Asp Asp Arg Asn Leu Ala Asp Met		
5 225	230	235
His Leu Gly Leu Tyr Asn Asp Val Ile Val Phe Asp His Val Glu Lys		
245	250	255
Lys Val Tyr Val Ile His Trp Val Arg Leu Asn Gln Gln Ser Ser Glu		
260	265	270
10 Glu Lys Ala Tyr Ala Glu Gly Leu Glu His Leu Glu Arg Leu Val Ser		
275	280	285
Arg Val Gln Asp Glu Asn Thr Pro Arg Leu Ala Pro Gly Ser Ile Asp		
290	295	300
Leu His Thr Gly His Phe Gly Pro Pro Leu Lys Lys Ser Asn Met Thr		
15 305	310	315
Cys Glu Glu Tyr Lys Met Ala Val Leu Ala Ala Lys Glu His Ile Gln		
325	330	335
Ala Gly Asp Ile Phe Gln Ile Val Leu Ser Gln Arg Phe Glu Arg Arg		
340	345	350
20 Thr Phe Ala Asp Pro Phe Glu Val Tyr Arg Ala Leu Arg Val Val Asn		
355	360	365
Pro Ser Pro Tyr Met Thr Tyr Met Gln Ala Arg Gly Cys Val Leu Val		
370	375	380
Ala Ser Ser Pro Glu Ile Leu Thr Arg Val Lys Lys Asn Lys Ile Val		
25 385	390	395
Asn Arg Pro Leu Ala Gly Thr Ala Arg Arg Gly Arg Thr Thr Glu Glu		
405	410	415
Asp Glu Met Leu Glu Thr Gln Leu Leu Lys Asp Ala Lys Gln Cys Ala		
420	425	430
30 Glu His Val Met Leu Val Asp Leu Gly Arg Asn Asp Val Gly Lys Val		
435	440	445
Ser Lys Ser Gly Ser Val Lys Val Glu Lys Leu Met Asn Val Glu Arg		
450	455	460
Tyr Ser His Val Met His Ile Ser Ser Thr Val Thr Gly Glu Leu Gln		
35 465	470	475
Asp Asn Leu Ser Cys Trp Asp Ala Leu Arg Ala Ala Leu Pro Val Gly		
485	490	495
Thr Val Ser Gly Ala Pro Lys Val Lys Ala Met Glu Leu Ile Asp Glu		
500	505	510
40 Leu Glu Val Asn Arg Arg Gly Pro Tyr Ser Gly Gly Phe Gly Gly Ile		

11

515	520	525
Ser Phe Thr Gly Asp Met Asp Ile Ala Leu Ala Leu Arg Thr Ile Val		
530	535	540
Phe Gln Thr Gly Thr Arg Tyr Asp Thr Met Tyr Ser Tyr Lys Asn Ala		
545	550	555
Thr Lys Arg Arg Gln Trp Val Ala Tyr Leu Gln Ala Gly Ala Gly Ile		560
565	570	575
Val Ala Asp Ser Asp Pro Asp Asp Glu His Arg Glu Cys Gln Asn Lys		
580	585	590
10 Ala Ala Gly Leu Ala Arg Ala Ile Asp Leu Ala Glu Ser Ala Phe Val		
595	600	605
Asn Lys Ser Ser Ser		
610		

15 <210> 7
 <211> 729
 <212> PRT
 <213> Rhizobium meliloti

20 <400> 7

Met Ala Ala Val Ile Leu Glu Asp Gly Ala Glu Ser Tyr Thr Thr Lys			
1	5	10	15
Gly Gly Ile Val Val Thr Arg Arg Arg Glu Ala Ser Tyr Ser Asp			
20	25	30	
25 Ala Ile Ala Gly Tyr Val Asp Arg Leu Asp Glu Arg Arg Gly Ala Val			
35	40	45	
Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr			
50	55	60	
Ala Val Val Asp Pro Pro Leu Ala Ile Ser Ser Phe Gly Arg Ser Leu			
30 65	70	75	80
Trp Ile Glu Ala Tyr Asn Glu Arg Gly Glu Val Leu Leu Ala Leu Ile			
85	90	95	
Ala Glu Asp Leu Lys Ser Val Ala Asp Ile Thr Leu Gly Ser Leu Ala			
100	105	110	
35 Ala Arg Arg Leu Asp Leu Thr Ile Asn Glu Pro Asp Arg Val Phe Thr			
115	120	125	
Glu Glu Glu Arg Ser Lys Met Pro Thr Val Phe Thr Val Leu Arg Ala			
130	135	140	
Val Thr Asn Leu Phe His Ser Glu Glu Asp Ser Asn Leu Gly Leu Tyr			
40 145	150	155	160

12

Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Glu Leu
 165 170 175
 Lys Leu Ser Arg Pro Asp Asp Gln Arg Asp Met Val Leu Phe Leu Pro
 180 185 190
 5 Asp Glu Ile Leu Val Val Asp His Tyr Ala Ala Lys Ala Trp Ile Asp
 195 200 205
 Arg Tyr Asp Phe Ala Arg Glu Asn Leu Ser Thr Glu Gly Lys Ala Ala
 210 215 220
 Asp Ile Ala Pro Glu Pro Phe Arg Ser Val Asp Ser Ile Pro Pro His
 10 225 230 235 240
 Gly Asp His Arg Pro Gly Glu Tyr Ala Glu Leu Val Val Lys Ala Lys
 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 15 Phe Tyr Glu Arg Cys Glu Ser Arg Pro Ser Glu Ile Ser Asn Arg Leu
 275 280 285
 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asn
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 20 305 310 315 320
 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Asn Ser
 340 345 350
 25 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 355 360 365
 Asp Lys Ser Arg Val Cys Val Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 30 385 390 395 400
 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 35 Arg Phe Ile Glu Ser His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 40 465 470 475 480

13

Ala Thr Leu Leu Tyr Asp Ser Asn Pro Glu Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ala Ala Ile Arg Asp Ala Lys Ser
 500 505 510
 5 Ala Asn Ser Ala Lys Ser Ala Arg Asp Val Ala Ala Val Gly Ala Gly
 515 520 525
 Val Ser Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Ser Val Thr Thr Val Arg Thr
 10 545 550 555 560
 Pro Val Ala Glu Glu Ile Phe Asp Arg Val Lys Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Thr Pro Lys Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 15 Ile Lys Lys Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Asp Leu Arg Gln Leu
 610 615 620
 Ala Ile Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 20 625 630 635 640
 Gly Ile Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ser Asn Leu Pro Arg Glu Phe Val Ile
 660 665 670
 25 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ser Lys
 675 680 685
 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 Gly Gly Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Ala His Leu
 30 705 710 715 720
 Ala Lys Arg Ala Lys Thr Lys Ala Ala
 725

<210> 8

35 <211> 421

<212> PRT

<213> Sulfolobus solfataricus

<400> 8

40 Met Glu Val His Pro Ile Ser Glu Phe Ala Ser Pro Phe Glu Val Phe

14

1	5	10	15
Lys Cys Ile Glu Arg Asp Phe Lys Val Ala Gly Leu Leu Glu Ser Ile			
20	25	30	
Gly Gly Pro Gln Tyr Lys Ala Arg Tyr Ser Val Ile Ala Trp Ser Thr			
5	35	40	45
Asn Gly Tyr Leu Lys Ile His Asp Asp Pro Val Asn Ile Leu Asn Gly			
50	55	60	
Tyr Leu Lys Asp Leu Lys Leu Ala Asp Ile Pro Gly Leu Phe Lys Gly			
65	70	75	80
10 Gly Met Ile Gly Tyr Ile Ser Tyr Asp Ala Val Arg Phe Trp Glu Lys			
85	90	95	
Ile Arg Asp Leu Lys Pro Ala Ala Glu Asp Trp Pro Tyr Ala Glu Phe			
100	105	110	
Phe Thr Pro Asp Asn Ile Ile Ile Tyr Asp His Asn Glu Gly Lys Val			
15	115	120	125
Tyr Val Asn Ala Asp Leu Ser Ser Val Gly Gly Cys Gly Asp Ile Gly			
130	135	140	
Glu Phe Lys Val Ser Phe Tyr Asp Glu Ser Leu Asn Lys Asn Ser Tyr			
145	150	155	160
20 Glu Arg Ile Val Ser Glu Ser Leu Glu Tyr Ile Arg Ser Gly Tyr Ile			
165	170	175	
Phe Gln Val Val Leu Ser Arg Phe Tyr Arg Tyr Ile Phe Ser Gly Asp			
180	185	190	
Pro Leu Arg Ile Tyr Tyr Asn Leu Arg Arg Ile Asn Pro Ser Pro Tyr			
25	195	200	205
Met Phe Tyr Leu Lys Phe Asp Glu Lys Tyr Leu Ile Gly Ser Ser Pro			
210	215	220	
Glu Leu Leu Phe Arg Val Gln Asp Asn Ile Val Glu Thr Tyr Pro Ile			
225	230	235	240
30 Ala Gly Thr Arg Pro Arg Gly Ala Asp Gln Glu Glu Asp Leu Lys Leu			
245	250	255	
Glu Leu Glu Leu Met Asn Ser Glu Lys Asp Lys Ala Glu His Leu Met			
260	265	270	
Leu Val Asp Leu Ala Arg Asn Asp Leu Gly Lys Val Cys Val Pro Gly			
35	275	280	285
Thr Val Lys Val Pro Glu Leu Met Tyr Val Glu Lys Tyr Ser His Val			
290	295	300	
Gln His Ile Val Ser Lys Val Ile Gly Thr Leu Lys Lys Tyr Asn			
305	310	315	320
40 Ala Leu Asn Val Leu Ser Ala Thr Phe Pro Ala Gly Thr Val Ser Gly			

15
325 330 335
Arg Pro Lys Pro Met Ala Met Asn Ile Ile Glu Thr Leu Glu Glu Tyr
340 345 350
Lys Arg Gly Pro Tyr Ala Gly Ala Val Gly Phe Ile Ser Ala Asp Gly
5 355 360 365
Asn Ala Glu Phe Ala Ile Ala Ile Arg Thr Ala Phe Leu Asn Lys Glu
370 375 380
Leu Leu Arg Ile His Ala Gly Ala Gly Ile Val Tyr Asp Ser Asn Pro
385 390 395 400
10 Glu Ser Glu Tyr Phe Glu Thr Glu His Lys Leu Lys Ala Leu Lys Thr
405 410 415
Ala Ile Gly Val Arg
420

15 <210> 9
<211> 32
<212> DNA
<213> Artificial Sequence

20 <220>
<223> A primer.

<400> 9
ccatcgccgc gcgtttttt cgtccaaacta tg

32

25 30
<210> 10
<211> 32
<212> DNA
<213> Artificial Sequence

30 35
<220>
<223> A primer.

<400> 10
35 catagttgga cgaaaaaaac ggcggcgat gg

32

<210> 11
<211> 39
<212> DNA
40 <213> Artificial Sequence

<220>
<223> A primer.

5 <400> 11
ccatcgccgc gcgatatttt cgtccaaacta tgaatatcc 39

<210> 12
<211> 39
10 <212> DNA
<213> Artificial Sequence

<220>
<223> A primer.
15
<400> 12
ggatattcat agttggacga aaaatacgcg ccgcgatgg 39

<210> 13
20 <211> 39
<212> DNA
<213> Artificial Sequence

<220>
25 <223> A primer.

<400> 13
ccatcgccgc gcgatatttt cgtccaaacta tgaatatcc 39

30 <210> 14
<211> 39
<212> DNA
<213> Artificial Sequence

35 <220>
<223> A primer.

<400> 14
ggatattcat agttggacga aaaccacgcg ccgcgatgg 39

17

<210> 15
<211> 39
<212> DNA
<213> Artificial Sequence

5

<220>
<223> A primer.

<400> 15
10 ccatcgccgc ggggtttta agtccaaacta tgaatatcc

39

<210> 16
<211> 39
<212> DNA
15 <213> Artificial Sequence

<220>
<223> A primer.

20 <400> 16
ggatattcat agttggactt aaaaaccgcg ccgcgatgg

39

<210> 17
<211> 34
25 <212> DNA
<213> Artificial Sequence

<220>
<223> A primer.

30
<400> 17
gccccgtttt ttctgtcaac tatgaatac cggg

34

<210> 18
35 <211> 34
<212> DNA
<213> Artificial Sequence

<220>
40 <223> A primer.

<400> 18
ccggatatt catagttgca cgaaaaaacc gcgc 34

5 <210> 19
<211> 34
<212> DNA
<213> Artificial Sequence

10 <220>
<223> A primer.

<400> 19
cgcggttttt tcgttcaact atgaatatcc gggc 34
15
<210> 20
<211> 34
<212> DNA
<213> Artificial Sequence

20
<220>
<223> A primer.

<400> 20
25 gcccgatata tcatagttga acgaaaaaac cgcg 34
<210> 21
<211> 37
<212> DNA
30 <213> Artificial Sequence

<220>
<223> A primer.

35 <400> 21
37 cggcgcggtt ttttcgatca actatgaata tccggc 37
<210> 22
<211> 37
40 <212> DNA

19

<213> Artificial Sequence

<220>

<223> A primer.

5

<400> 22

gccccggatat tcatacggtga tcgaaaaaac cgcgccg

37

<210> 23

10 <211> 36

<212> DNA

<213> Artificial Sequence

<220>

15 <223> A primer.

<400> 23

ggcgcgggttt ttcgctcaa ctatgaatat ccgggc

36

20 <210> 24

<211> 36

<212> DNA

<213> Artificial Sequence

25 <220>

<223> A primer.

<400> 24

gccccggatat tcatacggtga gcgaaaaaac cgcgcc

36

30

<210> 25

<211> 39

<212> DNA

<213> Artificial Sequence

35

<220>

<223> A primer.

<400> 25

40 cggcgcgggtt ttttcgatga actatgaata tccggggccg

39

<210> 26
<211> 39
<212> DNA
5 <213> Artificial Sequence

<220>
<223> A primer.

10 <400> 26
cgccccggat attccatagtt catcgaaaaa accgcgccc 39

<210> 27
<211> 34
15 <212> DNA
<213> Artificial Sequence

<220>
<223> A primer.

20
<400> 27
cgcgggtttt tcgaccaact atgaatatcc gggc 34

<210> 28
25 <211> 34
<212> DNA
<213> Artificial Sequence

<220>
30 <223> A primer.

<400> 28
gccccggatat tcatagttgg tcgaaaaaaaaac cgcg 34

35 <210> 29
<211> 36
<212> DNA
<213> Artificial Sequence

40 <220>

21

<223> A primer.

<400> 29

ggcgcggtt ttcggtaaa ctatgaatat ccgggc

36

5

<210> 30

<211> 36

<212> DNA

<213> Artificial Sequence

10

<220>

<223> A primer.

<400> 30

15 gccccgatat tcatagttga cggaaaaaac cgccgc

36

<210> 31

<211> 35

<212> DNA

20 <213> Artificial Sequence

<220>

<223> A primer.

25 <400> 31

gcgcggtttt ttctgtacaac tatgaatatac ccggcc

35

<210> 32

<211> 35

30 <212> DNA

<213> Artificial Sequence

<220>

<223> A primer.

35

<400> 32

gccccgatat tcatagttgt acggaaaaaac cgccgc

35

<210> 33

40 <211> 36

22

<212> DNA
<213> Artificial Sequence

<220>
5 <223> A primer.

<400> 33
cggcgccggtt ttttcgtcct tctatgaata tccggg 36

10 <210> 34
<211> 36
<212> DNA
<213> Artificial Sequence

15 <220>
<223> A primer.

<400> 34
cccgatatt catagaagga cgaaaaaacc ggcggc 36

20 <210> 35
<211> 29
<212> DNA
<213> Artificial Sequence

25 <220>
<223> A primer.

<400> 35
30 ctgaaggcga tcaacgcgtc gccctattc 29

<210> 36
<211> 29
<212> DNA
35 <213> Artificial Sequence

<220>
<223> A primer.

40 <400> 36

23

gaataggggcg acgcgtttagt cgccttcag

29

<210> 37

<211> 31

5 <212> DNA

<213> Artificial Sequence

<220>

<223> A primer.

10

<400> 37

cctgaaggcg atcaacgggt cgcccttattc c

31

<210> 38

15 <211> 31

<212> DNA

<213> Artificial Sequence

<220>

20 <223> A primer.

<400> 38

ggaaataggcg gaccggttga tgcgcattcag g

31

25 <210> 39

<211> 33

<212> DNA

<213> Artificial Sequence

30 <220>

<223> A primer.

<400> 39

cgtcgcctta ttccgccttc atcaatctcg gcg

33

35

<210> 40

<211> 33

<212> DNA

<213> Artificial Sequence

40

24

<220>

<223> A primer.

40

5 cgccgaggat gatgaaggcg gaatagggcg acg

33

41

33

DNA

10 <213> Artificial Sequence

42

<223> A primer.

41

cgtcgcccta ttctctggttc atcaaatctcg gcg

33

42

33

20 <212> DNA

<213> Artificial Sequence

43

<223> A primer.

25

42

cgccgaggat gatgaaccag gaatagggcg acg

33

43

30 <211> 729

<212> PRT

<213> Rhizobium meliloti

43

35 Met Ala Ala Val Ile Leu Glu Asp Gly Ala Glu Ser Tyr Thr Thr Lys

1

5

10

15

Gly Gly Ile Val Val Thr Arg Arg Arg Arg Glu Ala Ser Tyr Ser Asp

20

25

30

Ala Ile Ala Gly Tyr Val Asp Arg Leu Asp Glu Arg Arg Gly Ala Val

40

35

40

45

25

Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
 50 55 60
 Ala Val Val Asp Pro Pro Leu Ala Ile Ser Ser Phe Gly Arg Ser Leu
 65 70 75 80
 5 Trp Ile Glu Ala Tyr Asn Glu Arg Gly Glu Val Leu Leu Ala Leu Ile
 85 90 95
 Ala Glu Asp Leu Lys Ser Val Ala Asp Ile Thr Leu Gly Ser Leu Ala
 100 105 110
 Ala Arg Arg Leu Asp Leu Thr Ile Asn Glu Pro Asp Arg Val Phe Thr
 10 115 120 125
 Glu Glu Glu Arg Ser Lys Met Pro Thr Val Phe Thr Val Leu Arg Ala
 130 135 140
 Val Thr Asn Leu Phe His Ser Glu Glu Asp Ser Asn Leu Gly Leu Tyr
 145 150 155 160
 15 Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Glu Leu
 165 170 175
 Lys Leu Ser Arg Pro Asp Asp Gln Arg Asp Met Val Leu Phe Leu Pro
 180 185 190
 Asp Glu Ile Leu Val Val Asp His Tyr Ala Ala Lys Ala Trp Ile Asp
 20 195 200 205
 Arg Tyr Asp Phe Ala Arg Glu Asn Leu Ser Thr Glu Gly Lys Ala Ala
 210 215 220
 Asp Ile Ala Pro Glu Pro Phe Arg Ser Val Asp Ser Ile Pro Pro His
 225 230 235 240
 25 Gly Asp His Arg Pro Gly Glu Tyr Ala Glu Leu Val Val Lys Ala Lys
 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 Phe Tyr Glu Arg Cys Glu Ser Arg Pro Ser Glu Ile Ser Asn Arg Leu
 30 275 280 285
 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asn
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 35 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 40 355 360 365

Asp Lys Ser Arg Val Cys Val Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 5 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 Arg Phe Ile Glu Ser His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 10 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 15 Ala Thr Leu Leu Tyr Asp Ser Asn Pro Glu Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ala Ala Ile Arg Asp Ala Lys Ser
 500 505 510
 Ala Asn Ser Ala Lys Ser Ala Arg Asp Val Ala Ala Val Gly Ala Gly
 20 515 520 525
 Val Ser Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Ser Val Thr Thr Val Arg Thr
 545 550 555 560
 25 Pro Val Ala Glu Glu Ile Phe Asp Arg Val Lys Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Thr Pro Lys Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 Ile Lys Lys Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 30 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Asp Leu Arg Gln Leu
 610 615 620
 Ala Ile Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 35 Gly Ile Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ser Asn Leu Pro Arg Glu Phe Val Ile
 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ser Lys
 40 675 680 685

27

Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 Gly Gly Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Ala His Leu
 705 710 715 720
 5 Ala Lys Arg Ala Lys Thr Lys Ala Ala
 725

<210> 44
<211> 616
10 <212> PRT
<213> *Sulfolobus solfataricus*

<400> 44
Met Glu Val His Pro Ile Ser Glu Phe Ala Ser Pro Phe Glu Val Phe
15 1 5 10 15
Lys Cys Ile Glu Arg Asp Phe Lys Val Ala Gly Leu Leu Glu Ser Ile
 20 25 30
Gly Gly Pro Gln Tyr Lys Ala Arg Tyr Ser Val Ile Ala Trp Ser Thr
 35 40 45
20 Asn Gly Tyr Leu Lys Ile His Asp Asp Pro Val Asn Ile Leu Asn Gly
 50 55 60
Tyr Leu Lys Asp Leu Lys Leu Ala Asp Ile Pro Gly Leu Phe Lys Gly
 65 70 75 80
Gly Met Ile Gly Tyr Ile Ser Tyr Asp Ala Val Arg Phe Trp Glu Lys
25 85 90 95
Ile Arg Asp Leu Lys Pro Ala Ala Glu Asp Trp Pro Tyr Ala Glu Phe
 100 105 110
Phe Thr Pro Asp Asn Ile Ile Tyr Asp His Asn Glu Gly Lys Val
 115 120 125
30 Tyr Val Asn Ala Asp Leu Ser Ser Val Gly Gly Cys Gly Asp Ile Gly
 130 135 140
Glu Phe Lys Val Ser Phe Tyr Asp Glu Ser Leu Asn Lys Asn Ser Tyr
 145 150 155 160
Glu Arg Ile Val Ser Glu Ser Leu Glu Tyr Ile Arg Ser Gly Tyr Ile
35 165 170 175
Phe Gln Val Val Leu Ser Arg Phe Tyr Arg Tyr Ile Phe Ser Gly Asp
 180 185 190
Pro Leu Arg Ile Tyr Tyr Asn Leu Arg Arg Ile Asn Pro Ser Pro Tyr
 195 200 205
40 Met Phe Tyr Leu Lys Phe Asp Glu Lys Tyr Leu Ile Gly Ser Ser Pro

28

210	215	220
Glu Leu Leu Phe Arg Val Gln Asp Asn Ile Val Glu Thr Tyr Pro Ile		
225	230	235
Ala Gly Thr Arg Pro Arg Gly Ala Asp Gln Glu Glu Asp Leu Lys Leu		240
5	245	250
Glu Leu Glu Leu Met Asn Ser Glu Lys Asp Lys Ala Glu His Leu Met		255
260	265	270
Leu Val Asp Leu Ala Arg Asn Asp Leu Gly Lys Val Cys Val Pro Gly		
275	280	285
10 Thr Val Lys Val Pro Glu Leu Met Tyr Val Glu Lys Tyr Ser His Val		
290	295	300
Gln His Ile Val Ser Lys Val Ile Gly Thr Leu Lys Lys Lys Tyr Asn		
305	310	315
Ala Leu Asn Val Leu Ser Ala Thr Phe Pro Ala Gly Thr Val Ser Gly		320
15	325	330
Arg Pro Lys Pro Met Ala Met Asn Ile Ile Glu Thr Leu Glu Glu Tyr		335
340	345	350
Lys Arg Gly Pro Tyr Ala Gly Ala Val Gly Phe Ile Ser Ala Asp Gly		
355	360	365
20 Asn Ala Glu Phe Ala Ile Ala Ile Arg Thr Ala Phe Leu Asn Lys Glu		
370	375	380
Leu Leu Arg Ile His Ala Gly Ala Gly Ile Val Tyr Asp Ser Asn Pro		
385	390	395
Glu Ser Glu Tyr Phe Glu Thr Glu His Lys Leu Lys Ala Leu Lys Thr		400
25	405	410
Ala Ile Gly Val Arg Met Asp Leu Thr Leu Ile Ile Asp Asn Tyr Asp		415
420	425	430
Ser Phe Val Tyr Asn Ile Ala Gln Ile Val Gly Glu Leu Gly Ser Tyr		
435	440	445
30 Pro Ile Val Ile Arg Asn Asp Glu Ile Ser Ile Lys Gly Ile Glu Arg		
450	455	460
Ile Asp Pro Asp Arg Leu Ile Ile Ser Pro Gly Pro Gly Thr Pro Glu		
465	470	475
Lys Arg Glu Asp Ile Gly Val Ser Leu Asp Val Ile Lys Tyr Leu Gly		480
35	485	490
Lys Arg Thr Pro Ile Leu Gly Val Cys Leu Gly His Gln Ala Ile Gly		495
500	505	510
Tyr Ala Phe Gly Ala Lys Ile Arg Arg Ala Arg Lys Val Phe His Gly		
515	520	525
40 Lys Ile Ser Asn Ile Ile Leu Val Asn Asn Ser Pro Leu Ser Leu Tyr		

29

530	535	540
Tyr Gly Ile Ala Lys Glu Phe Lys Ala Thr Arg Tyr His Ser Leu Val		
545	550	555
Val Asp Glu Val His Arg Pro Leu Ile Val Asp Ala Ile Ser Ala Glu		560
5	565	570
Asp Asn Glu Ile Met Ala Ile His His Glu Glu Tyr Pro Ile Tyr Gly		575
580	585	590
Val Gln Phe His Pro Glu Ser Val Gly Thr Ser Leu Gly Tyr Lys Ile		
595	600	605
10 Leu Tyr Asn Phe Leu Asn Arg Val		
610	615	

<210> 45			
<211> 897			
15 <212> PRT			
<213> Arabidopsis thaliana			
<400> 45			
Met Ser Ala Val Ser Ile Ser Ala Val Lys Ser Asp Phe Phe Thr Val			
20 1	5	10	15
Glu Ala Ile Ala Val Thr His His Arg Thr Pro His Pro Pro His Phe			
20	25	30	
Pro Ser Leu Arg Phe Pro Leu Ser Leu Lys Ser Pro Pro Ala Thr Ser			
35	40	45	
25 Leu Asn Leu Val Ala Gly Ser Lys Leu Leu His Phe Ser Arg Arg Leu			
50	55	60	
Pro Ser Ile Lys Cys Ser Tyr Thr Pro Ser Leu Asp Leu Ser Glu Glu			
65	70	75	80
Gln Phe Thr Lys Phe Lys Lys Ala Ser Glu Lys Gly Asn Leu Val Pro			
30	85	90	95
Leu Phe Arg Cys Val Phe Ser Asp His Leu Thr Pro Ile Leu Ala Tyr			
100	105	110	
Arg Cys Leu Val Lys Glu Asp Asp Arg Asp Ala Pro Ser Phe Leu Phe			
115	120	125	
35 Glu Ser Val Glu Pro Gly Ser Gln Ser Ser Asn Ile Gly Arg Tyr Ser			
130	135	140	
Val Val Gly Ala Gln Pro Thr Ile Glu Ile Val Ala Lys Gly Asn Val			
145	150	155	160
Val Thr Val Met Asp His Gly Ala Ser Leu Arg Thr Glu Glu Glu Val			
40	165	170	175

30

Asp	Asp	Pro	Met	Met	Val	Pro	Gln	Lys	Ile	Met	Glu	Glu	Trp	Asn	Pro			
			180				185								190			
Gln	Gly	Ile	Asp	Glu	Leu	Pro	Glu	Ala	Phe	Cys	Gly	Gly	Trp	Val	Gly			
			195				200								205			
5	Tyr	Phe	Ser	Tyr	Asp	Thr	Val	Arg	Tyr	Val	Glu	Lys	Lys	Lys	Leu	Pro		
			210				215								220			
Phe	Ser	Asn	Ala	Pro	Glu	Asp	Asp	Arg	Ser	Leu	Pro	Asp	Val	Asn	Leu			
			225				230				235				240			
Gly	Leu	Tyr	Asp	Asp	Val	Ile	Val	Phe	Asp	His	Val	Glu	Lys	Lys	Ala			
10			245					250							255			
Tyr	Val	Ile	His	Trp	Val	Arg	Ile	Asp	Lys	Asp	Arg	Ser	Val	Glu	Glu			
								265							270			
Asn	Phe	Arg	Glu	Gly	Met	Asn	Arg	Leu	Glu	Ser	Leu	Thr	Ser	Arg	Ile			
			275				280				285							
15	Gln	Asp	Gln	Lys	Pro	Pro	Lys	Met	Pro	Thr	Gly	Phe	Ile	Lys	Leu	Arg		
								290			295				300			
Thr	Gln	Leu	Phe	Gly	Pro	Lys	Leu	Glu	Lys	Ser	Thr	Met	Thr	Ser	Glu			
								305		310		315			320			
			Ala	Tyr	Lys	Glu	Ala	Val	Val	Glu	Ala	Lys	Glu	His	Ile	Leu	Ala	Gly
20								325			330				335			
Asp	Ile	Phe	Gln	Ile	Val	Leu	Ser	Gln	Arg	Phe	Glu	Arg	Arg	Thr	Phe			
								340		345					350			
Ala	Asp	Pro	Phe	Glu	Ile	Tyr	Arg	Ala	Leu	Arg	Ile	Val	Asn	Pro	Ser			
								355		360					365			
25	Pro	Tyr	Met	Ala	Tyr	Leu	Gln	Val	Arg	Gly	Cys	Ile	Leu	Val	Ala	Ser		
								370		375					380			
Ser	Pro	Glu	Ile	Leu	Leu	Arg	Ser	Lys	Asn	Arg	Lys	Ile	Thr	Asn	Arg			
								385		390		395			400			
			Pro	Leu	Ala	Gly	Thr	Val	Arg	Arg	Gly	Lys	Thr	Pro	Lys	Glu	Asp	Leu
30								405			410				415			
Met	Leu	Glu	Glu	Leu	Leu	Ser	Asp	Glu	Lys	Gln	Cys	Ala	Glu	His				
								420		425		430						
			Ile	Met	Leu	Val	Asp	Leu	Gly	Arg	Asn	Asp	Val	Gly	Lys	Val	Ser	Lys
								435		440		445						
35	Pro	Gly	Ser	Val	Glu	Val	Lys	Lys	Leu	Lys	Asp	Ile	Glu	Trp	Phe	Ser		
								450		455					460			
His	Val	Met	His	Ile	Ser	Ser	Thr	Val	Val	Gly	Glu	Leu	Leu	Asp	His			
								465		470		475			480			
			Leu	Thr	Ser	Trp	Asp	Ala	Leu	Arg	Ala	Val	Leu	Pro	Val	Gly	Thr	Val
40								485			490				495			

31

Ser Gly Ala Pro Lys Val Lys Ala Met Glu Leu Ile Asp Glu Leu Glu
 500 505 510
 Val Thr Arg Arg Gly Pro Tyr Ser Gly Gly Phe Gly Gly Ile Ser Phe
 515 520 525
 5 Asn Gly Asp Met Asp Ile Ala Leu Ala Leu Arg Thr Met Val Phe Pro
 530 535 540
 Thr Asn Thr Arg Tyr Asp Thr Leu Tyr Ser Tyr Lys His Pro Gln Arg
 545 550 555 560
 Arg Arg Glu Trp Ile Ala His Ile Gln Ala Gly Ala Gly Ile Val Ala
 10 565 570 575
 Asp Ser Asn Pro Asp Asp Glu His Arg Glu Cys Glu Asn Lys Ala Ala
 580 585 590
 Ala Leu Ala Arg Ala Ile Asp Leu Ala Glu Ser Ser Phe Leu Glu Ala
 595 600 605
 15 Pro Glu Phe Thr Thr Ile Thr Pro His Ile Asn Asn Ile Met Ala Ala
 610 615 620
 Ser Thr Leu Tyr Lys Ser Cys Leu Leu Gln Pro Lys Ser Gly Ser Thr
 625 630 635 640
 Thr Arg Arg Leu Asn Pro Ser Leu Val Asn Pro Leu Thr Asn Pro Thr
 20 645 650 655
 Arg Val Ser Val Leu Gly Lys Ser Arg Arg Asp Val Phe Ala Lys Ala
 660 665 670
 Ser Ile Glu Met Ala Glu Ser Asn Ser Ile Pro Ser Val Val Asn
 675 680 685
 25 Ser Ser Lys Gln His Gly Pro Ile Ile Val Ile Asp Asn Tyr Asp Ser
 690 695 700
 Phe Thr Tyr Asn Leu Cys Gln Tyr Met Gly Glu Leu Gly Cys His Phe
 705 710 715 720
 Glu Val Tyr Arg Asn Asp Glu Leu Thr Val Glu Glu Leu Lys Lys Lys
 30 725 730 735
 Asn Pro Arg Gly Val Leu Ile Ser Pro Gly Pro Gly Thr Pro Gln Asp
 740 745 750
 Ser Gly Ile Ser Leu Gln Thr Val Leu Glu Leu Gly Pro Leu Val Pro
 755 760 765
 35 Leu Phe Gly Val Cys Met Gly Leu Gln Cys Ile Gly Glu Ala Phe Gly
 770 775 780
 Gly Lys Ile Val Arg Ser Pro Phe Gly Val Met His Gly Lys Ser Ser
 785 790 795 800
 Met Val His Tyr Asp Glu Lys Gly Glu Gly Leu Phe Ser Gly Leu
 40 805 810 815

32

Ser Asn Pro Phe Ile Val Gly Arg Tyr His Ser Leu Val Ile Glu Lys			
820	825	830	
Asp Thr Phe Pro Ser Asp Glu Leu Glu Val Thr Ala Trp Thr Glu Asp			
835	840	845	
5 Gly Leu Val Met Ala Ala Arg His Arg Lys Tyr Lys His Ile Gln Gly			
850	855	860	
Val Gln Phe His Pro Glu Ser Ile Ile Thr Thr Glu Gly Lys Thr Ile			
865	870	875	880
Val Arg Asn Phe Ile Lys Ile Val Glu Lys Lys Glu Ser Glu Lys Leu			
10	885	890	895
Thr			

<210> 46			
15 <211> 252			
<212> DNA			
<213> Artificial Sequence			
 20 <220>			
<223> A truncated gene			
 <400> 46			
atgcaaac acaaaaaccgac tctcgaactg gaattccctgg tggaaaacgg tatcgccacc	60		
gtgcgaacgg gtgtctgggt agtccttcatatgtatccgc agtcggaaagg cgacgaaacc	120		
25 cgttaacaaag cccgcgtgt actgcgcgtt attgcacccg cgcattatgc acaggagact	180		
tctgtatggc tgacattctg ctgctcgata atatcgactc ttttacgtac aacctggcag	240		
atcagttgcg ca	252		
 <210> 47			
30 <211> 18			
<212> DNA			
<213> Artificial Sequence			
 <220>			
35 <223> A primer.			
 <400> 47			
ttatgccggcc tgtcatcg	18		
 40 <210> 48			

33

<211> 19
<212> DNA
<213> Artificial Sequence

5 <220>
<223> A primer.

<400> 48
ataggcttaa tggttaaccg

19

10
<210> 49
<211> 18
<212> DNA
<213> Artificial Sequence

15
<220>
<223> A primer.

<400> 49
20 ctgaacaaca gaagtacg

18

<210> 50
<211> 18
<212> DNA
25 <213> Artificial Sequence

<220>
<223> A primer.

30 <400> 50
taaccgtgtc atcgagcg

18

<210> 51
<211> 31
35 <212> DNA
<213> Artificial Sequence

<220>
<223> A primer

40

34

<400> 51
aaaaagatct ccatggtaac gatcattcag g 31

<210> 52
5 <211> 35
<212> DNA
<213> Artificial Sequence

<220>
10 <223> A primer

<400> 52
aaaagaattc ttatcacgca gccttggct tcgcc 35

15 <210> 53
<211> 19
<212> DNA
<213> Artificial Sequence

20 <220>
<223> A primer

<400> 53
caaaaagctgg atccccacc 19
25
<210> 54
<211> 23
<212> DNA
<213> Artificial Sequence

30
<220>
<223> A primer

<400> 54
35 cctatccgag atctctcaac tcc 23

<210> 55
<211> 31
<212> DNA
40 <213> Artificial Sequence

<220>
<223> A primer

5 <400> 55
catcccatgg atggtaacga tcattcagga t 31

<210> 56
<211> 31
10 <212> DNA
<213> Artificial Sequence

<220>
<223> A primer
15
<400> 56
gatgtctaga gacactatag aatactcaag c 31

<210> 57
20 <211> 719
<212> PRT
<213> Rhodopseudomonas palustris

<400> 57

25 Met Asn Arg Thr Val Phe Ser Leu Pro Ala Thr Ser Asp Tyr Lys Thr
1 5 10 15
Ala Ala Gly Leu Ala Val Thr Arg Ser Ala Gln Pro Phe Ala Gly Gly
20 25 30
Gln Ala Leu Asp Glu Leu Ile Asp Leu Leu Asp His Arg Arg Gly Val
30 35 40 45
Met Leu Ser Ser Gly Thr Thr Val Pro Gly Arg Tyr Glu Ser Phe Asp
50 55 60
Leu Gly Phe Ala Asp Pro Pro Leu Ala Leu Thr Thr Arg Ala Glu Lys
65 70 75 80
35 Phe Thr Ile Glu Ala Leu Asn Pro Arg Gly Arg Val Leu Ile Ala Phe
85 90 95
Leu Ser Asp Lys Leu Glu Glu Pro Cys Val Val Glu Gln Ala Cys
100 105 110
Ala Thr Lys Ile Arg Gly His Ile Val Arg Gly Glu Ala Pro Val Asp
40 115 120 125

Glu Glu Gln Arg Thr Arg Arg Ala Ser Ala Ile Ser Leu Val Arg Ala
 130 135 140
 Val Ile Ala Ala Phe Ala Ser Pro Ala Asp Pro Met Leu Gly Leu Tyr
 145 150 155 160
 5 Gly Ala Phe Ala Tyr Asp Leu Val Phe Gln Phe Glu Asp Leu Lys Gln
 165 170 175
 Lys Arg Ala Arg Glu Ala Asp Gln Arg Asp Ile Val Leu Tyr Val Pro
 180 185 190
 Asp Arg Leu Leu Ala Tyr Asp Arg Ala Thr Gly Arg Gly Val Asp Ile
 10 195 200 205
 Ser Tyr Glu Phe Ala Trp Lys Gly Gln Ser Thr Ala Gly Leu Pro Asn
 210 215 220
 Glu Thr Ala Glu Ser Val Tyr Thr Gln Thr Gly Arg Gln Gly Phe Ala
 225 230 235 240
 15 Asp His Ala Pro Gly Asp Tyr Pro Lys Val Val Glu Lys Ala Arg Ala
 245 250 255
 Ala Phe Ala Arg Gly Asp Leu Phe Glu Ala Val Pro Gly Gln Leu Phe
 260 265 270
 Gly Glu Pro Cys Glu Arg Ser Pro Ala Glu Val Phe Lys Arg Leu Cys
 20 275 280 285
 Arg Ile Asn Pro Ser Pro Tyr Gly Gly Leu Leu Asn Leu Gly Asp Gly
 290 295 300
 Glu Phe Leu Val Ser Ala Ser Pro Glu Met Phe Val Arg Ser Asp Gly
 305 310 315 320
 25 Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Ala Arg Gly Val
 325 330 335
 Asp Ala Ile Ser Asp Ala Glu Gln Ile Gln Lys Leu Leu Asn Ser Glu
 340 345 350
 Lys Asp Glu Phe Glu Leu Asn Met Cys Thr Asp Val Asp Arg Asn Asp
 30 355 360 365
 Lys Ala Arg Val Cys Val Pro Gly Thr Ile Lys Val Leu Ala Arg Arg
 370 375 380
 Gln Ile Glu Thr Tyr Ser Lys Leu Phe His Thr Val Asp His Val Glu
 385 390 395 400
 35 Gly Met Leu Arg Pro Gly Phe Asp Ala Leu Asp Ala Phe Leu Thr His
 405 410 415
 Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met Gln
 420 425 430
 Phe Val Glu Asp His Glu Arg Ser Pro Arg Arg Trp Tyr Ala Gly Ala
 40 435 440 445

37

Phe Gly Val Val Gly Phe Asp Gly Ser Ile Asn Thr Gly Leu Thr Ile
 450 455 460
 Arg Thr Ile Arg Met Lys Asp Gly Leu Ala Glu Val Arg Val Gly Ala
 465 470 475 480
 5 Thr Cys Leu Phe Asp Ser Asn Pro Val Ala Glu Asp Lys Glu Cys Gln
 485 490 495
 Val Lys Ala Ala Ala Leu Phe Gln Ala Leu Arg Gly Asp Pro Ala Lys
 500 505 510
 Pro Leu Ser Ala Val Ala Pro Asp Ala Thr Gly Ser Gly Lys Lys Val
 10 515 520 525
 Leu Leu Val Asp His Asp Asp Ser Phe Val His Met Leu Ala Asp Tyr
 530 535 540
 Phe Arg Gln Val Gly Ala Gln Val Thr Val Val Arg Tyr Val His Gly
 545 550 555 560
 15 Leu Lys Met Leu Ala Glu Asn Ser Tyr Asp Leu Leu Val Leu Ser Pro
 565 570 575
 Gly Pro Gly Arg Pro Glu Asp Phe Lys Ile Lys Asp Thr Ile Asp Ala
 580 585 590
 Ala Leu Ala Lys Lys Leu Pro Ile Phe Gly Val Cys Leu Gly Val Gln
 20 595 600 605
 Ala Met Gly Glu Tyr Phe Gly Gly Thr Leu Gly Gln Leu Ala Gln Pro
 610 615 620
 Ala His Gly Arg Pro Ser Arg Ile Gln Val Arg Gly Gly Ala Leu Met
 625 630 635 640
 25 Arg Gly Leu Pro Asn Glu Val Thr Ile Gly Arg Tyr His Ser Leu Tyr
 645 650 655
 Val Asp Met Arg Asp Met Pro Lys Glu Leu Thr Val Thr Ala Ser Thr
 660 665 670
 Asp Asp Gly Ile Ala Met Ala Ile Glu His Lys Thr Leu Pro Val Gly
 30 675 680 685
 Gly Val Gln Phe His Pro Glu Ser Leu Met Ser Leu Gly Gly Glu Val
 690 695 700
 Gly Leu Arg Ile Val Glu Asn Ala Phe Arg Leu Gly Gln Ala Ala
 705 710 715

35
 <210> 58
 <211> 729
 <212> PRT
 <213> Artificial Sequence

<220>

<223> An A. tumefaciens mutant.

<400> 58

5	Met	Val	Thr	Ile	Ile	Gln	Asp	Asp	Gly	Ala	Glu	Thr	Tyr	Glu	Thr	Lys		
1																15		
	10																	
		Gly	Gly	Ile	Ile	Gln	Val	Ser	Arg	Lys	Arg	Arg	Pro	Thr	Asp	Tyr	Ala	Asn
	20																30	
		Ala	Ile	Asp	Asn	Tyr	Ile	Glu	Lys	Leu	Asp	Ser	His	Arg	Gly	Ala	Phe	
10	35															45		
		Phe	Ser	Ser	Asn	Tyr	Glu	Tyr	Pro	Gly	Arg	Tyr	Thr	Arg	Trp	Asp	Thr	
	50															55	60	
		Ala	Ile	Val	Asp	Pro	Pro	Leu	Gly	Ile	Ser	Cys	Phe	Gly	Arg	Lys	Met	
	65															70	75	80
15	Trp	Ile	Glu	Ala	Tyr	Asn	Gly	Arg	Gly	Glu	Val	Leu	Leu	Asp	Phe	Ile		
																85	90	95
		Thr	Glu	Lys	Leu	Lys	Ala	Thr	Pro	Asp	Leu	Thr	Leu	Gly	Ala	Ser	Ser	
																100	105	110
		Thr	Arg	Arg	Leu	Asp	Leu	Thr	Val	Asn	Glu	Pro	Asp	Arg	Val	Phe	Thr	
20	115															120	125	
		Glu	Glu	Glu	Arg	Ser	Lys	Ile	Pro	Thr	Val	Phe	Thr	Ala	Leu	Arg	Ala	
	130															135	140	
		Ile	Val	Asp	Leu	Phe	Tyr	Ser	Ser	Ala	Asp	Ser	Ala	Ile	Gly	Leu	Phe	
	145															150	155	160
		25	Gly	Ala	Phe	Gly	Tyr	Asp	Leu	Ala	Phe	Gln	Phe	Asp	Ala	Ile	Lys	Leu
																165	170	175
		Ser	Leu	Ala	Arg	Pro	Glu	Asp	Gln	Arg	Asp	Met	Val	Leu	Phe	Leu	Pro	
																180	185	190
		Asp	Glu	Ile	Leu	Val	Val	Asp	His	Tyr	Ser	Ala	Lys	Ala	Trp	Ile	Asp	
30	195															200	205	
		Arg	Tyr	Asp	Phe	Glu	Lys	Asp	Gly	Met	Thr	Thr	Asp	Gly	Lys	Ser	Ser	
	210															215	220	
		Asp	Ile	Thr	Pro	Asp	Pro	Phe	Lys	Thr	Thr	Asp	Thr	Ile	Pro	Pro	Lys	
	225															230	235	240
		35	Gly	Asp	His	Arg	Pro	Gly	Glu	Tyr	Ser	Glu	Leu	Val	Val	Lys	Ala	Lys
																245	250	255
		Glu	Ser	Phe	Arg	Arg	Gly	Asp	Leu	Phe	Glu	Val	Val	Pro	Gly	Gln	Lys	
																260	265	270
		Phe	Met	Glu	Arg	Cys	Glu	Ser	Asn	Pro	Ser	Ala	Ile	Ser	Arg	Arg	Leu	
40	275															280	285	

39

Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 5 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 10 355 360 365
 Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 15 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 20 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 25 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 500 505 510
 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 30 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 35 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 40 595 600 605

40

Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
610 615 620
Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
625 630 635 640
5 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
645 650 655
His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
660 665 670
Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
10 675 680 685
Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
690 695 700
Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
705 710 715 720
15 Thr Arg Lys Ala Lys Thr Lys Ala Ala
725

<210> 59
<211> 729
20 <212> PRT
<213> Artificial Sequence

<220>
<223> An A. tumefaciens mutant.
25
<400> 59
Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys
1 5 10 15
Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
30 20 25 30
Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Tyr
35 40 45
Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
50 55 60
35 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
65 70 75 80
Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile
85 90 95
Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
40 100 105 110

41

Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
 115 120 125
 Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
 130 135 140
 5 Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
 145 150 155 160
 Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
 165 170 175
 Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
 10 180 185 190
 Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp
 195 200 205
 Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser
 210 215 220
 15 Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys
 225 230 235 240
 Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys
 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 20 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 275 280 285
 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 290 295 300
 25 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 30 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 355 360 365
 Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 35 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 40 420 425 430

42

Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 5 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 10 500 505 510
 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 15 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 20 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 25 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 30 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 675 680 685
 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 35 Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
 705 710 715 720
 Thr Arg Lys Ala Lys Thr Lys Ala Ala
 725

43

<211> 729

<212> PRT

<213> Artificial Sequence

5 <220>

<223> An A. tumefaciens mutant.

<400> 60

Met	Val	Thr	Ile	Ile	Gln	Asp	Asp	Gly	Ala	Glu	Thr	Tyr	Glu	Thr	Lys	
10	1			5					10				15			
Gly	Gly	Ile	Gln	Val	Ser	Arg	Lys	Arg	Arg	Pro	Thr	Asp	Tyr	Ala	Asn	
		20						25				30				
Ala	Ile	Asp	Asn	Tyr	Ile	Glu	Lys	Leu	Asp	Ser	His	Arg	Gly	Ala	Val	
		35				40				45						
15	Phe	Ser	Phe	Asn	Tyr	Glu	Tyr	Pro	Gly	Arg	Tyr	Thr	Arg	Trp	Asp	Thr
	50					55			60							
Ala	Ile	Val	Asp	Pro	Pro	Leu	Gly	Ile	Ser	Cys	Phe	Gly	Arg	Lys	Met	
	65					70			75			80				
Trp	Ile	Glu	Ala	Tyr	Asn	Gly	Arg	Gly	Glu	Val	Leu	Leu	Asp	Phe	Ile	
20					85			90			95					
Thr	Glu	Lys	Leu	Lys	Ala	Thr	Pro	Asp	Leu	Thr	Leu	Gly	Ala	Ser	Ser	
		100					105			110						
Thr	Arg	Arg	Leu	Asp	Leu	Thr	Val	Asn	Glu	Pro	Asp	Arg	Val	Phe	Thr	
	115					120				125						
25	Glu	Glu	Glu	Arg	Ser	Lys	Ile	Pro	Thr	Val	Phe	Thr	Ala	Leu	Arg	Ala
	130					135			140							
Ile	Val	Asp	Leu	Phe	Tyr	Ser	Ser	Ala	Asp	Ser	Ala	Ile	Gly	Leu	Phe	
	145					150			155			160				
Gly	Ala	Phe	Gly	Tyr	Asp	Leu	Ala	Phe	Gln	Phe	Asp	Ala	Ile	Lys	Leu	
30					165			170			175					
Ser	Leu	Ala	Arg	Pro	Glu	Asp	Gln	Arg	Asp	Met	Val	Leu	Phe	Leu	Pro	
					180			185			190					
Asp	Glu	Ile	Leu	Val	Val	Asp	His	Tyr	Ser	Ala	Lys	Ala	Trp	Ile	Asp	
		195				200			205							
35	Arg	Tyr	Asp	Phe	Glu	Lys	Asp	Gly	Met	Thr	Thr	Asp	Gly	Lys	Ser	Ser
	210					215			220							
Asp	Ile	Thr	Pro	Asp	Pro	Phe	Lys	Thr	Thr	Asp	Thr	Ile	Pro	Pro	Lys	
	225					230			235			240				
Gly	Asp	His	Arg	Pro	Gly	Glu	Tyr	Ser	Glu	Leu	Val	Val	Lys	Ala	Lys	
40					245			250			255					

44

Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 275 280 285
 5 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 10 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 355 360 365
 15 Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 20 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 435 440 445
 25 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Glu Ala Glu Thr
 30 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 500 505 510
 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 515 520 525
 35 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 40 565 570 575

45

Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605
 5 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 10 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 675 680 685
 15 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
 705 710 715 720
 Thr Arg Lys Ala Lys Thr Lys Ala Ala
 20 725

<210> 61

<211> 729

<212> PRT

25 <213> Artificial Sequence

<220>

<223> An A. tumefaciens mutant.

30 <400> 61
 Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys
 1 5 10 15
 Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
 20 25 30
 35 Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val
 35 40 45
 Phe Ser Cys Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
 50 55 60
 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
 40 65 70 75 80

46

Trp Ile Glu Ala Tyr Asn Gly Arg Gly	Glu Val Leu Leu Asp Phe Ile	
85	90	95
Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser		
100	105	110
5 Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr		
115	120	125
Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala		
130	135	140
Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe		
10 145	150	155
Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu		
165	170	175
Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro		
180	185	190
15 Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp		
195	200	205
Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser		
210	215	220
Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys		
20 225	230	235
Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys		
245	250	255
Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys		
260	265	270
25 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu		
275	280	285
Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp		
290	295	300
Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser		
30 305	310	315
Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly		
325	330	335
Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser		
340	345	350
35 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn		
355	360	365
Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg		
370	375	380
Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile		
40 385	390	395
		400

47

Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 5 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 10 465 470 475 480
 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 500 505 510
 15 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 20 545 550 555 560
 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 25 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 30 625 630 635 640
 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 660 665 670
 35 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 675 680 685
 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
 40 705 710 715 720

Thr Arg Lys Ala Lys Thr Lys Ala Ala
725

<210> 62
5 <211> 729
<212> PRT
<213> Artificial Sequence

<220>
10 <223> An A. tumefaciens mutant.

<400> 62
Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys
1 5 10 15
15 Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
20 25 30
Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val
35 40 45
Phe Ser Ser Phe Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
20 50 55 60
Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
65 70 75 80
Trp Ile Glu Ala Tyr Asn Gly Arg Gly Val Leu Leu Asp Phe Ile
85 90 95
25 Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
100 105 110
Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
115 120 125
Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
30 130 135 140
Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
145 150 155 160
Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
165 170 175
35 Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
180 185 190
Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp
195 200 205
Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser
40 210 215 220

49

Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys
 225 230 235 240
 Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys
 245 250 255
 5 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 275 280 285
 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 10 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 15 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 355 360 365
 Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 20 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 25 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 30 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Ala Glu Thr
 485 490 495
 35 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 500 505 510
 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 40 530 535 540

50

Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560

Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575

5 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 580 585 590

Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605

Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 10 610 615 620

Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640

Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655

15 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 660 665 670

Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 675 680 685

Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 20 690 695 700

Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
 705 710 715 720

Thr Arg Lys Ala Lys Thr Lys Ala Ala
 725

25

<210> 63
 <211> 729
 <212> PRT
 <213> Artificial Sequence

30

<220>
 <223> An A. tumefaciens mutant.

<400> 63

35 Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys
 1 5 10 15

Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
 20 25 30

Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val
 40 35 40 45

51

Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
 50 55 60
 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
 65 70 75 80
 5 Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile
 85 90 95
 Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
 100 105 110
 Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
 10 115 120 125
 Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
 130 135 140
 Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
 145 150 155 160
 15 Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
 165 170 175
 Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
 180 185 190
 Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp
 20 195 200 205
 Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser
 210 215 220
 Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys
 225 230 235 240
 25 Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys
 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 30 275 280 285
 Lys Ala Ile Asn Ala Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 35 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 40 355 360 365

52

Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 5 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 10 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 15 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 500 505 510
 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 20 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 25 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 30 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 35 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 40 675 680 685

53

Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
690 695 700
Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
705 710 715 720
5 Thr Arg Lys Ala Lys Thr Lys Ala Ala
725

<210> 64

<211> 729

10 <212> PRT

<213> Artificial Sequence

<220>

<223> An A. tumefaciens mutant.

15

<400> 64

Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys
1 5 10 15
Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
20 20 25 30
Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val
35 40 45
Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
50 55 60
25 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
65 70 75 80
Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile
85 90 95
Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
30 100 105 110
Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
115 120 125
Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
130 135 140
35 Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
145 150 155 160
Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
165 170 175
Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
40 180 185 190

54

Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp
 195 200 205
 Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser
 210 215 220
 5 Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys
 225 230 235 240
 Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys
 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 10 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 275 280 285
 Lys Ala Ile Asn Gly Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 290 295 300
 15 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser
 20 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 355 360 365
 Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 25 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 30 420 425 430
 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 35 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 40 500 505 510

55

Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 5 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 10 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 15 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 20 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 675 680 685
 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 25 Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val His Leu
 705 710 715 720
 Thr Arg Lys Ala Lys Thr Lys Ala Ala
 725

30 <2110> 65

<2111> 729

<2112> PRT

<2113> Artificial Sequence

35 <2200>

<223> An A. tumefaciens mutant.

<400> 65

Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys

40 1

5

10

15

56

Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
 20 25 30
 Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val
 35 40 45
 5 Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
 50 55 60
 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
 65 70 75 80
 Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile
 10 85 90 95
 Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
 100 105 110
 Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
 115 120 125
 15 Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
 130 135 140
 Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
 145 150 155 160
 Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
 20 165 170 175
 Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
 180 185 190
 Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp
 195 200 205
 25 Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser
 210 215 220
 Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys
 225 230 235 240
 Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys
 30 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 275 280 285
 35 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Trp Phe Ile Asn Leu Gly Asp
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 40 325 330 335

57

Asp	Asp	Pro	Ile	Ala	Asp	Ser	Glu	Gln	Ile	Leu	Lys	Leu	Leu	Asn	Ser		
			340				345								350		
Lys	Lys	Asp	Glu	Ser	Glu	Leu	Thr	Met	Cys	Ser	Asp	Val	Asp	Arg	Asn		
			355				360								365		
5	Asp	Lys	Ser	Arg	Val	Cys	Glu	Pro	Gly	Ser	Val	Lys	Val	Ile	Gly	Arg	
			370				375								380		
Arg	Gln	Ile	Glu	Met	Tyr	Ser	Arg	Leu	Ile	His	Thr	Val	Asp	His	Ile		
			385				390								395		400
10	Glu	Gly	Arg	Leu	Arg	Asp	Asp	Met	Asp	Ala	Phe	Asp	Gly	Phe	Leu	Ser	
			405												410		415
His	Ala	Trp	Ala	Val	Thr	Val	Thr	Gly	Ala	Pro	Lys	Leu	Trp	Ala	Met		
			420												425		430
Arg	Phe	Ile	Glu	Gly	His	Glu	Lys	Ser	Pro	Arg	Ala	Trp	Tyr	Gly	Gly		
			435												440		445
15	Ala	Ile	Gly	Met	Val	Gly	Phe	Asn	Gly	Asp	Met	Asn	Thr	Gly	Leu	Thr	
			450				455								460		
Leu	Arg	Thr	Ile	Arg	Ile	Lys	Asp	Gly	Ile	Ala	Glu	Val	Arg	Ala	Gly		
			465				470								475		480
20	Ala	Thr	Leu	Leu	Asn	Asp	Ser	Asn	Pro	Gln	Glu	Glu	Ala	Glu	Thr		
			485												490		495
Glu	Leu	Lys	Ala	Ser	Ala	Met	Ile	Ser	Ala	Ile	Arg	Asp	Ala	Lys	Gly		
			500												505		510
Thr	Asn	Ser	Ala	Ala	Ala	Thr	Lys	Arg	Asp	Ala	Ala	Lys	Val	Gly	Thr	Gly	
			515												520		525
25	Val	Lys	Ile	Leu	Leu	Val	Asp	His	Glu	Asp	Ser	Phe	Val	His	Thr	Leu	
			530				535								540		
Ala	Asn	Tyr	Phe	Arg	Gln	Thr	Gly	Ala	Thr	Val	Ser	Thr	Val	Arg	Ser		
			545				550								555		560
30	Pro	Val	Ala	Ala	Asp	Val	Phe	Asp	Arg	Phe	Gln	Pro	Asp	Leu	Val	Val	
			565												570		575
Leu	Ser	Pro	Gly	Pro	Gly	Ser	Pro	Thr	Asp	Phe	Asp	Cys	Lys	Ala	Thr		
			580												585		590
Ile	Lys	Ala	Ala	Arg	Ala	Arg	Asp	Leu	Pro	Ile	Phe	Gly	Val	Cys	Leu		
			595												600		605
35	Gly	Leu	Gln	Ala	Leu	Ala	Glu	Ala	Tyr	Gly	Gly	Glu	Leu	Arg	Gln	Leu	
			610				615								620		
Ala	Val	Pro	Met	His	Gly	Lys	Pro	Ser	Arg	Ile	Arg	Val	Leu	Glu	Pro		
			625				630								635		640
40	Gly	Leu	Val	Phe	Ser	Gly	Leu	Gly	Lys	Glu	Val	Thr	Val	Gly	Arg	Tyr	
			645												650		655

58

His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
660 665 670
Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
675 680 685
5 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
690 695 700
Gly Cln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
705 710 715 720
Thr Arg Lys Ala Lys Thr Lys Ala Ala
10 725

<210> 66
<211> 604
<212> PRT
15 <213> Artificial Sequence

<220>
<223> A Zea mays mutant.

20 <400> 66
Met Glu Ser Leu Ala Ala Thr Ser Val Phe Ala Pro Ser Arg Val Ala
1 5 10 15
Val Pro Ala Ala Arg Ala Leu Val Arg Ala Gly Thr Val Val Pro Thr
20 25 30
25 Arg Arg Thr Ser Ser Arg Ser Gly Thr Ser Gly Val Lys Cys Ser Ala
35 40 45
Ala Val Thr Pro Gln Ala Ser Pro Val Ile Ser Arg Ser Ala Ala Ala
50 55 60
Ala Lys Ala Ala Glu Glu Asp Lys Arg Arg Phe Phe Glu Ala Ala Ala
30 65 70 75 80
Arg Gly Ser Gly Lys Gly Asn Leu Val Pro Met Trp Glu Cys Ile Val
85 90 95
Ser Asp His Leu Thr Pro Val Leu Ala Tyr Arg Cys Leu Val Pro Glu
100 105 110
35 Asp Asn Val Asp Ala Pro Ser Phe Leu Phe Glu Ser Val Glu Gln Gly
115 120 125
Pro Gln Gly Thr Thr Asn Val Gly Arg Tyr Ser Met Val Gly Ala His
130 135 140
40 Pro Val Met Glu Ile Val Ala Lys Asp His Lys Val Thr Ile Met Asp
145 150 155 160

59

His	Glu	Lys	Ser	Gln	Val	Thr	Glu	Gln	Val	Val	Asp	Asp	Pro	Met	Gln	
					165				170					175		
Ile	Pro	Arg	Thr	Met	Met	Glu	Gly	Trp	His	Pro	Gln	Gln	Ile	Asp	Glu	
					180				185				190			
5	Leu	Pro	Glu	Ser	Phe	Ser	Gly	Gly	Trp	Val	Gly	Phe	Phe	Ser	Tyr	Asp
					195				200			205				
Thr	Val	Arg	Tyr	Val	Glu	Lys	Lys	Leu	Pro	Phe	Ser	Ser	Ala	Pro		
					210				215		220					
Gln	Asp	Asp	Arg	Asn	Leu	Pro	Asp	Val	His	Leu	Gly	Leu	Tyr	Asp	Asp	
10	225					230				235			240			
Val	Leu	Val	Phe	Asp	Asn	Val	Glu	Lys	Lys	Val	Tyr	Val	Ile	His	Trp	
					245				250			255				
Val	Asn	Val	Asp	Arg	His	Ala	Ser	Val	Glu	Glu	Ala	Tyr	Gln	Asp	Gly	
					260				265			270				
15	Arg	Ser	Arg	Leu	Asn	Met	Leu	Leu	Ser	Lys	Val	His	Asn	Ser	Asn	Val
					275				280			285				
Pro	Thr	Leu	Ser	Pro	Gly	Phe	Val	Lys	Leu	His	Thr	Arg	Lys	Phe	Gly	
					290				295			300				
Thr	Pro	Leu	Asn	Lys	Ser	Thr	Met	Thr	Ser	Asp	Glu	Tyr	Lys	Asn	Ala	
20	305					310				315			320			
Val	Leu	Gln	Ala	Lys	Glu	His	Ile	Met	Ala	Gly	Asp	Ile	Phe	Gln	Ile	
					325				330			335				
Val	Leu	Ser	Gln	Arg	Phe	Glu	Arg	Arg	Thr	Tyr	Ala	Asn	Pro	Phe	Glu	
					340				345			350				
25	Val	Tyr	Arg	Ala	Leu	Arg	Ile	Val	Asn	Pro	Ser	Pro	Tyr	Lys	Ala	Tyr
					355				360			365				
Val	Gln	Ala	Arg	Gly	Cys	Val	Leu	Val	Ala	Ser	Ser	Pro	Glu	Ile	Leu	
					370				375			380				
Thr	Arg	Val	Ser	Lys	Gly	Lys	Ile	Ile	Asn	Arg	Pro	Leu	Ala	Gly	Thr	
30	385					390				395			400			
Val	Arg	Arg	Gly	Lys	Thr	Glu	Lys	Glu	Asp	Gln	Met	Gln	Glu	Gln	Gln	
					405				410			415				
Leu	Leu	Ser	Asp	Glu	Lys	Gln	Cys	Ala	Glu	His	Ile	Met	Leu	Val	Asp	
					420				425			430				
35	Leu	Gly	Arg	Asn	Asp	Val	Gly	Lys	Val	Ser	Lys	Pro	Gly	Ser	Val	Lys
						435			440			445				
Val	Glu	Lys	Leu	Met	Asn	Ile	Glu	Arg	Tyr	Ser	His	Val	Met	His	Ile	
					450				455			460				
Ser	Ser	Thr	Val	Ser	Gly	Gln	Leu	Asp	Asp	His	Leu	Gln	Ser	Trp	Asp	
40	465					470				475			480			

60

Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly Ala Pro Lys
 485 490 495
 Val Lys Ala Met Glu Leu Ile Asp Lys Leu Glu Val Thr Arg Arg Gly
 500 505 510
 5 Pro Tyr Ser Gly Gly Leu Gly Ile Ser Phe Asp Gly Asp Met Gln
 515 520 525
 Ile Ala Leu Ser Leu Arg Thr Ile Val Phe Ser Thr Ala Pro Ser His
 530 535 540
 Asn Thr Met Tyr Ser Tyr Lys Asp Ala Asp Arg Arg Arg Glu Trp Val
 10 545 550 555 560
 Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ser Pro Asp
 565 570 575
 Asp Glu Gln Arg Glu Cys Glu Asn Lys Ala Ala Ala Leu Ala Arg Ala
 580 585 590
 15 Ile Asp Leu Ala Glu Ser Ala Phe Val Asp Lys Glu
 595 600

<210> 67

<211> 1815

20 <212> DNA

<213> Artificial Sequence

<220>

<223> A Zea mays mutant.

25

<400> 67

atggaaatccc tagccgcac	ctccgtgttc	gcccctccc	gctgcgcgt	ccggggggcg	60
cggccctgg	ttagggcggg	gacggtgta	ccaaccaggc	ggacgagcag	120
accagcgggg	tgaatgtctc	tgctggcg	aegccgcagg	cgagccccagt	180
30 agcgctgccc	cgccgaaaggc	ggcgaggagg	gacaaggaggc	ggttttcg	240
cgggggagcc	ggaaggggaa	cctgggtccc	atgtggggat	gcategtgtc	300
accccccgtc	tegcctactcg	ctgcctcg	cccgaggaca	acgtcgacgc	360
ctcttcgagt	ccgtcgagca	ggggccccag	ggcacccaca	acgtcgcccg	420
gtgggagccc	acccagtgt	ggagattgt	gccaaaggacc	ctatagcatg	480
35 cacgagaaga	gccaagtgtac	agagcaggta	gtggcagacc	cccgaggacc	540
atgatgggg	gatggcaccc	acagacatc	gacgacgtcc	ctgaatctt	600
tgggttgggt	tcttttctta	tgatacggtt	aggtatgtt	agaagaagaa	660
tccagtgtc	ctcaggacga	taggaacctt	cctgtatgtc	acttgggact	720
40 cggcatgtcat	ctgttgagga	agcataccaa	gatggcagg	ctatgtggac	780
			cccgactaaa	catgttgc	840

61

tctaaaagtgc acaattccaa	tgtccccaca ctctcttcgt	gatttgtcaa gtcgcacaca	900
cgcggatttt gtacaccctt	gaacaagtcg accatgacaa	gtgatgatgt taagaatgt	960
gttctgcagg ctaaggaaca	tattatggct ggggatatac	tccagatgtt ttaagccag	1020
agggtcgaga	gacgaacata tgccaaacca	tttgagggtt atcgagcatt acggatttg	1080
5 aactcttagcc	cataacaaggc gtatgtacag	gcaagggct gtgttattgt tgctgtttagt	1140
cctggaaatcc ttacacgagt	cagtaagggg aagattatta	atcgaccact tgctggaaat	1200
gttctgaaggc gcaagacaga	gaaggaaatgt caaatgcag	agcagcaact gttaaatgtat	1260
gaaaaaactgt	gtggcggaca cataatgttt	gtggacttgg gaaggaaatgt tggtggcaag	1320
gtatccaaac caggatcaat	gaagggtggg aagttgtatgt	acattggagag atactcccat	1380
10 gttatgcacaa tcagtcacac	ggtttagtggaa cagttggatg	atcatcttca gagttggat	1440
gccttggagat ctgccttgcc	ctgtggaaaca gtcagttgg	cacccaaatgtt gaaggccat	1500
gagttgttggat ataaatgtt	tttttttttttggat	tttttttttttggat	1560
atatacgatcc ttgtgttgc	atgtgttgcat	gcataattgc ctttcttcg	1620
gcgcgcgagcc acaacacat	gtactctatac	aaagacgcag ataggcgat	1680
15 gctcatcttc aggctgtgtc	aggcattgtt	ggccgacatgt gcccagatgt	1740
gaatgcgaga ataaggctgc	tgcaactatgt	cggggcatatgt atcttgcaga	1800
gtgacacaa aataag		gtcagatgtt	1815

<210> 68

20 <211> 2204

<212> DNA

<213> Artificial Sequence

<220>

25 <223> A Zea mays mutant.

<400> 68

atggaaatccc tagccgcac	ctcccggttgc	gcccctccc gctgcgcgt	ccggggggcg	60	
cggggcccttgg	tttagggcggtt	gacgggtgtt	ccaaccaggc	ggacgacgcg	120
30 accagggggg	tgaatgtctc	tgctggccgt	acggccgcgg	cgagcccaatgt	180
agcgtcgccg	cgccgaaaggc	ggccggaggag	gacaagaggc	ggttttcg	240
cggggggagcc	ggaaaggggaa	cctgggttccc	atgtggggat	gcatcggttc	300
accccccgtc	tcgtctatcc	ctggctcttc	ccggaggaca	acgtcgacgc	360
cttttcgatgt	ccgtcgacgt	ggggccccag	ggcacccaca	acgtcgcccg	420
35 gtggggagccc	acccaggatgt	ggagattgt	gccaaagacc	acaagggttac	480
cacgagaaga	gccaaggatgc	agagcaggta	gtggacacc	cgatcgatgt	540
atgatgtggg	gatggccatcc	acagcagatc	gacggatcc	ctgaatcttc	600
tgggttgggt	tcttttccta	tgatcgtt	aggtatgtt	agaagaagaa	660
40 tcaatgtgtc	ctcaggacgt	taggaaccc	cctgtatgtc	acttggact	720
gttcttagtct	tcgtataatgt	tgagaagaaa	gtatgttta	tccattgggt	780

62

cggcatgcat	ctgttgagga	agcataccaa	gatggcagggt	cccgactaaa	catgttgcta	840
tctaaagtgc	acaattccaa	tgtccccaca	ctctcttcgt	gatttgcgaa	gctgcacaca	900
cgcaagttt	gtacacccctt	gaacaagtcg	accatgacaa	gtgtgagta	taagaatgt	960
gtttcgcagg	ctaaggaaaca	tattatggct	ggggatatct	tccagatgt	tttaaggccag	1020
5 aggttcgaga	gacgaacata	tgccaaccca	tttgagggtt	atcgagcatt	acggatttgt	1080
aatccctagcc	catacaaggc	gtatgtacag	gcaagggtt	gtgttattgt	tgcgtctagt	1140
cctgaaattc	ttiacacagt	cgtaaagggg	aagttattata	atcgaccact	tgctggaaact	1200
gttcgaaggg	gcaagacaga	gaaggaaagat	caaatgcgaa	agcagcaact	gttaatgtgt	1260
gaaaaaacagt	gtgcccggca	cataatgcctt	ttggacttgg	gaaggaaatgt	tgttggcaag	1320
10 gtatccaaac	caggatcgat	gaagggtggg	aagtgtatgt	acattggagag	atactcccat	1380
gttgcaca	tcagtcac	gggttagtgg	cagttggatg	atcatctcca	gagttgggat	1440
gccttgcgg	ctgccttgc	cgttggaaaca	gtcagtgtgg	cacccaaagggt	gaaggccatgt	1500
gagttgattt	ataaagtgg	agttacgggg	cgaggaccat	atagtgttgg	tctaggaggg	1560
atatcgttt	atgttgacat	gcaaaattgtca	ctttcttc	gcaccatcg	attctcaaca	1620
15 ggcgcgaccc	acaacaegat	gtactcatac	aaagacgcag	ataggcgtcg	ggagtgggtc	1680
gctcatcttc	aggctgtgc	aggcattgtt	gccgcacagt	gcccagatgt	cgaacaacgt	1740
gaatgcgaga	ataaggctgc	tgcaactagct	cgggccatcg	atcttgacaga	gtcagctttt	1800
gtagacaaag	ataatgtgc	tatggttatc	gttttagttct	tgttcatgtt	tcttttaccc	1860
actttccgtt	aaaaaaaaat	gtcatttagt	ggtgggaaaa	agcaataaga	ctgttctcta	1920
20 gaattcggc	tcggtaccgg	atccaattcc	cgatcgatca	aacatttggc	aataaagt	1980
cttaagattt	aatctgttgc	ccggcttgc	gatgtatc	atataatttc	tgttgaatta	2040
cgttaagcat	gtataatatta	acatgtatgt	catgcgtt	tttatgagat	gggtttttat	2100
gat tagatgc	ccgcattat	acatattata	cgcgataga	aaaaaaat	agcgcgcaaa	2160
cattggataaa	ttatcgccgc	cggtgtcata	tatgttacta	gatc		2204

25

<210> 69

<211> 729

<212> PRT

<213> Artificial Sequence

30

<220>

<223> An *A. tumefaciens* mutant.

<400> 69

35 Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys

1 5 10 15

Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn

20 25 30

Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val

40

35

40

45

63

Phe Lys Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
 50 55 60
 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
 65 70 75 80
 5 Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile
 85 90 95
 Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
 100 105 110
 Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
 10 115 120 125
 Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
 130 135 140
 Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
 145 150 155 160
 15 Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
 165 170 175
 Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
 180 185 190
 Asp Glu Ile Leu Val Val Asp His Tyr Ser Ala Lys Ala Trp Ile Asp
 20 195 200 205
 Arg Tyr Asp Phe Glu Lys Asp Gly Met Thr Thr Asp Gly Lys Ser Ser
 210 215 220
 Asp Ile Thr Pro Asp Pro Phe Lys Thr Thr Asp Thr Ile Pro Pro Lys
 225 230 235 240
 25 Gly Asp His Arg Pro Gly Glu Tyr Ser Glu Leu Val Val Lys Ala Lys
 245 250 255
 Glu Ser Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Lys
 260 265 270
 Phe Met Glu Arg Cys Glu Ser Asn Pro Ser Ala Ile Ser Arg Arg Leu
 30 275 280 285
 Lys Ala Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Asp
 290 295 300
 Gln Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Ser
 305 310 315 320
 35 Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly
 325 330 335
 Asp Asp Pro Ile Ala Asp Ser Glu Gln Ile Leu Lys Leu Asn Ser
 340 345 350
 Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn
 40 355 360 365

64

Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg
 370 375 380
 Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile
 385 390 395 400
 5 Glu Gly Arg Leu Arg Asp Asp Met Asp Ala Phe Asp Gly Phe Leu Ser
 405 410 415
 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met
 420 425 430
 Arg Phe Ile Glu Gly His Glu Lys Ser Pro Arg Ala Trp Tyr Gly Gly
 10 435 440 445
 Ala Ile Gly Met Val Gly Phe Asn Gly Asp Met Asn Thr Gly Leu Thr
 450 455 460
 Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu Val Arg Ala Gly
 465 470 475 480
 15 Ala Thr Leu Leu Asn Asp Ser Asn Pro Gln Glu Glu Ala Glu Thr
 485 490 495
 Glu Leu Lys Ala Ser Ala Met Ile Ser Ala Ile Arg Asp Ala Lys Gly
 500 505 510
 Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 20 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 25 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 30 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 35 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 40 675 680 685

65

Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
690 695 700
Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
705 710 715 720
5 Thr Arg Lys Ala Lys Thr Lys Ala Ala
725

<210> 70

<211> 729

10 <212> PRT

<213> Artificial Sequence

<220>

<223> An A. tumefaciens mutant.

15

<400> 70

Met Val Thr Ile Ile Gln Asp Asp Gly Ala Glu Thr Tyr Glu Thr Lys
1 5 10 15
Gly Gly Ile Gln Val Ser Arg Lys Arg Arg Pro Thr Asp Tyr Ala Asn
20 20 25 30
Ala Ile Asp Asn Tyr Ile Glu Lys Leu Asp Ser His Arg Gly Ala Val
35 40 45
Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr
50 55 60
25 Ala Ile Val Asp Pro Pro Leu Gly Ile Ser Cys Phe Gly Arg Lys Met
65 70 75 80
Trp Ile Glu Ala Tyr Asn Gly Arg Gly Glu Val Leu Leu Asp Phe Ile
85 90 95
Thr Glu Lys Leu Lys Ala Thr Pro Asp Leu Thr Leu Gly Ala Ser Ser
30 100 105 110
Thr Arg Arg Leu Asp Leu Thr Val Asn Glu Pro Asp Arg Val Phe Thr
115 120 125
Glu Glu Glu Arg Ser Lys Ile Pro Thr Val Phe Thr Ala Leu Arg Ala
130 135 140
35 Ile Val Asp Leu Phe Tyr Ser Ser Ala Asp Ser Ala Ile Gly Leu Phe
145 150 155 160
Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Ala Ile Lys Leu
165 170 175
Ser Leu Ala Arg Pro Glu Asp Gln Arg Asp Met Val Leu Phe Leu Pro
40 180 185 190

66

Asp	Glu	Ile	Leu	Val	Val	Asp	His	Tyr	Ser	Ala	Lys	Ala	Trp	Ile	Asp	
195						200					205					
Arg	Tyr	Asp	Phe	Glu	Lys	Asp	Gly	Met	Thr	Thr	Asp	Gly	Lys	Ser	Ser	
210						215					220					
5	Asp	Ile	Thr	Pro	Asp	Pro	Phe	Lys	Thr	Thr	Asp	Thr	Ile	Pro	Pro	Lys
225						230					235				240	
Gly	Asp	His	Arg	Pro	Gly	Glu	Tyr	Ser	Glu	Leu	Val	Val	Lys	Ala	Lys	
245						250					255					
Glu	Ser	Phe	Arg	Arg	Gly	Asp	Leu	Phe	Glu	Val	Val	Pro	Gly	Gln	Lys	
10						260					265			270		
Phe	Met	Glu	Arg	Cys	Glu	Ser	Asn	Pro	Ser	Ala	Ile	Ser	Arg	Arg	Leu	
275						280					285					
Lys	Ala	Ile	Asn	Pro	Ser	Pro	Tyr	Ser	Ala	Phe	Ile	Asn	Leu	Gly	Asp	
290						295					300					
15	Gln	Glu	Tyr	Leu	Val	Gly	Ala	Ser	Pro	Glu	Met	Phe	Val	Arg	Val	Ser
305						310					315				320	
Gly	Arg	Arg	Ile	Glu	Thr	Cys	Pro	Ile	Ser	Gly	Thr	Ile	Lys	Arg	Gly	
325						330					335					
Asp	Asp	Pro	Ile	Ala	Asp	Ser	Glu	Gln	Ile	Leu	Lys	Leu	Leu	Asn	Ser	
20						340					345			350		
Lys	Lys	Asp	Glu	Ser	Glu	Leu	Thr	Met	Cys	Ser	Asp	Val	Asp	Arg	Asn	
355						360					365					
Asp	Lys	Ser	Arg	Val	Cys	Glu	Pro	Gly	Ser	Val	Lys	Val	Ile	Gly	Arg	
370						375					380					
25	Arg	Gln	Ile	Glu	Met	Tyr	Ser	Arg	Leu	Ile	His	Thr	Val	Asp	His	Ile
385						390					395			400		
Glu	Gly	Arg	Leu	Arg	Asp	Asp	Met	Asp	Ala	Phe	Asp	Gly	Phe	Leu	Ser	
405						410					415					
His	Ala	Trp	Ala	Val	Thr	Val	Thr	Gly	Ala	Pro	Lys	Leu	Trp	Ala	Met	
30						420					425			430		
Arg	Phe	Ile	Glu	Gly	His	Glu	Lys	Ser	Pro	Arg	Ala	Trp	Tyr	Gly	Gly	
435						440					445					
Ala	Ile	Gly	Met	Val	Gly	Phe	Asn	Gly	Asp	Met	Asn	Thr	Gly	Leu	Thr	
450						455					460					
35	Leu	Arg	Thr	Ile	Arg	Ile	Lys	Asp	Gly	Ile	Ala	Glu	Val	Arg	Ala	Gly
465						470					475			480		
Ala	Thr	Leu	Leu	Asn	Asp	Ser	Asn	Pro	Gln	Glu	Glu	Ala	Glu	Thr		
485						490					495					
Glu	Leu	Lys	Ala	Ser	Ala	Met	Ile	Ser	Ala	Ile	Arg	Asp	Ala	Lys	Gly	
40						500					505			510		

67

Thr Asn Ser Ala Ala Thr Lys Arg Asp Ala Ala Lys Val Gly Thr Gly
 515 520 525
 Val Lys Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu
 530 535 540
 5 Ala Asn Tyr Phe Arg Gln Thr Gly Ala Thr Val Ser Thr Val Arg Ser
 545 550 555 560
 Pro Val Ala Ala Asp Val Phe Asp Arg Phe Gln Pro Asp Leu Val Val
 565 570 575
 Leu Ser Pro Gly Pro Gly Ser Pro Thr Asp Phe Asp Cys Lys Ala Thr
 10 580 585 590
 Ile Lys Ala Ala Arg Ala Arg Asp Leu Pro Ile Phe Gly Val Cys Leu
 595 600 605
 Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu Leu Arg Gln Leu
 610 615 620
 15 Ala Val Pro Met His Gly Lys Pro Ser Arg Ile Arg Val Leu Glu Pro
 625 630 635 640
 Gly Leu Val Phe Ser Gly Leu Gly Lys Glu Val Thr Val Gly Arg Tyr
 645 650 655
 His Ser Ile Phe Ala Asp Pro Ala Thr Leu Pro Arg Asp Phe Ile Ile
 20 660 665 670
 Thr Ala Glu Ser Glu Asp Gly Thr Ile Met Gly Ile Glu His Ala Lys
 675 680 685
 Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu
 690 695 700
 25 Gly Gln Asp Ala Gly Met Arg Met Ile Glu Asn Val Val Val His Leu
 705 710 715 720
 Thr Arg Lys Ala Lys Thr Lys Ala Ala
 725

30 <210> 71
 <211> 264
 <212> DNA
 <213> Artificial Sequence

35 <220>
 <223> The sequence of a CTP.

<400> 71
 atggcttcct ctaggtcttc ttccgctact atggttgcct ctccggctca ggcactatg 60
 40 gtcgcttcct tcaacggact taagtcctcc gtcgccttcc cagccacccg caaggctaac 120

68

aacgacatta cttccatcac aagcaacggc ggaagaggtta actgcatgca ggtgtggct	180
ccgattggaa agaagaaggtt tgagacttc tcttacacctc ctgacacctac cgattccggt	240
ggtcgcgtca actgcatgca ggcc	264

5 <210> 72
 <211> 88
 <212> PRT
 <213> Artificial Sequence

10 <220>
 <223> The sequence of a CTP.

<400> 72
 Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
 15 1 5 10 15
 Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
 20 30
 Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
 35 45
 20 Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Gly Lys
 50 60
 Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser Gly
 65 80
 Gly Arg Val Asn Cys Met Gln Ala
 25 85

<210> 73
 <211> 264
 <212> DNA
 30 <213> Artificial Sequence

<220>
 <223> The sequence of a CTP.

35 <400> 73
 atggcttcctt ctatgctctc ttccgctact atggttgcct ctccggctca ggccactatg 60
 gtcgctccctt tcaaaccggact taaggccctcc gctgccttc cagccaccccg caaggctaaac 120
 aacgacatta cttccatcac aagcaacggc ggaagaggtta actgcatgca ggtgtggct 180
 ccgattggaaa agaagaaggtt tgagacttc tcttacacctc ctgacacctac cgattccggt 240
 40 ggttcgcgtca actgcatgca ggcc 264

<210> 74
<211> 88
<212> PRT
5 <213> Artificial Sequence

<220>
<223> The sequence of a CTP.

10 <400> 74
Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
1 5 10 15
Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
20 25 30
15 Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
35 40 45
Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Glu Lys
50 55 60
Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser Gly
20 65 70 75 80
Gly Arg Val Asn Cys Met Gln Ala
85

<210> 75
25 <211> 2190
<212> DNA
<213> Artificial Sequence

<220>
30 <223> An optimized A. tumefaciens.

<400> 75	
atggtagacca tcattcagga tgacggtgcc gagacctacg agaccaaggg cggcatccag	60
gtgagccgca ageccgcgcc caccgattac gccaacgcca tcgataacta catcgaaaaag	120
35 ctgttattcccc atcgcgggtgc cgtgtttctcc tccaaactacg aatacccccagg ccgtcacacc	180
cgctgggata ccggccatcgat cgatccacca ctccgcatt cctgcttcgg ccgcggatgt	240
tggatcgaaag cctacaacgg ccgcggcggaa gtgcgtctcg atttcattac cgaaaagctg	300
aaggccacac ccgatctcac cctcggcgct tccctccaccc ggcgcctcga tcttacccgtc	360
aacgaaccac accgcgttcc caccgaagaa gaacgcgtccaa aatcccaac cgttccatcc	420
40 gctctcaggcc ccatcgatcgat cctttctac tccagccggcc attccggccat cggccgttcc	480

70

ggtgccttcg	gttacgatct	ccgccttcag	ttcgacggcca	tcaagcttcc	cctggccccgc	540	
ccagaagacc	agcgcgacat	ggtgctgttc	ctggcccgtat	aaatccctgt	cgttgtatcac	600	
tactccgcca	aggcctggat	cgaccgcatac	gatttcgaga	aggacggcat	gaccaccgac	660	
ggcaaattct	ccgcacattac	ccccgatccc	ttcaagacca	ccgataccat	cccacccaag	720	
5	ggcgcattacc	ccccccggca	atactccgag	cttgtgttgc	aggccaaaggaa	aegcttcggc	780
cgccgcgacc	tgttcggagt	cgttccggc	cagaaatcta	ttggagcgctg	cgaaaagcaac	840	
ccatccgcca	tttcccccgg	cctgaaggccc	atcaaccat	ccccctactc	cttcttcate	900	
aacctcggc	atcaggaata	cctgtcggt	gcttcggccag	aatgtttctgt	ggcggttcttc	960	
ggccgcgcca	tcgagacctg	cccaatctca	ggcaccatca	agcgcggcga	cgatccaat	1020	
10	ggccgcgacag	aggacgatttt	gaaactgtc	aactccaaaa	aggacgaaatc	cgaactgacc	1080
atgtgtccg	acgtggacccg	caacgacaag	agccgggtct	gcgcggcagg	ttccgtgaag	1140	
gtcattggcc	gcgcgcagat	cgagatgtac	tcacgctca	ttccacccgt	cgatcacatc	1200	
gaaggccggc	tgccgcgacga	tatggacccc	ttcgacgggtt	ttctcagccca	cgccctgggcc	1260	
gtcaccgtca	ccgggtgacc	aaagctgtgg	gccatcgctt	tcatcgaaagg	tcatgaaaag	1320	
15	agccccacgg	cctgtacgg	cggtgcctatc	ggcatgggt	gcttcaacgg	cgacatgaac	1380
acggcctgt	ccctgcgcac	catccgcata	aaggacggta	ttggccgaagt	gcgcgcggc	1440	
gcacccctgc	tcaacgattc	caaccacag	gaagaagaag	cgaaaacgca	actgaaggcc	1500	
tcggccatgt	tctcagccat	tcgcgcacga	aaaggccacca	actctggccg	caccaagggc	1560	
gtatggccca	aagtggccac	cggcgtcaag	atccgtgtcg	tcgaccacga	agacagcttc	1620	
20	gtgcacaccc	tgcccaacta	ttccggccag	accggccca	ccgtctccac	cgtcagggtca	1680
ccagtcgcag	ccgcacgttt	cgatcgctt	caggccgacc	tcgttgttct	gtccccccgt	1740	
ccggcggcc	caaccggattt	cgactcgaaag	gcacccatca	aggccgcggc	cgccggcgat	1800	
ctgcacatet	tccggcggtt	cctcggtct	caggcattgg	cagaagctca	cgggggcgag	1860	
ctgcgcggc	tttgcgtgtcc	catgcacggc	aaggcttccc	gcatccgggt	gttggaaaccc	1920	
25	ggctctgtct	tctccggttt	cggaaggaa	gtcacccgt	gtcgatccaa	ttccatottc	1980
ggcgatcccg	ccacccctgc	acgcgatttc	atcatcaccg	cagaagcgaa	ggacggccacc	2040	
atcatgggc	togaacacgc	caaggaaacc	gtggccgcgg	ttcagttcca	cccagaatcc	2100	
atcatgaccc	tccgtcgagg	cgccggcatg	cgcatgatcg	agaaacgtcg	ggtgcacatcg	2160	
acccgcgg	ccaaagacca	ggccgcctgt				2190	
30							
<2110>	76						
<211>	2160						
<212>	DNA						
<213>	Rhodopseudomonas palustris						
35							
<400>	76						
atgaacagga	ccgttttctc	gcttccgcg	accaggcact	ataagaccgc	cgccggccctc	60	
gcgggtacgc	gcagcgcaca	gccttttgc	ggccggccagg	cgctcgacga	gtgtatcgat	120	
ctgcgtacacc	accggccgggg	cgtgtatcg	tcgtccggca	caaccgtgc	ggccgcgtac	180	
40	gagagcttcg	acccctgggtt	tgccgatccg	ccgctggcgc	tcaccactag	ggccgaaaaa	240

	71	
ttcaccatcg aggccgtcaa tccgcgcgc cgggtgctga tcgcgttct gtccgacaag	300	
cttgaagagc cctcgcttgt ggtggagcag gcctgcgc ccaagatcg gggccacatc	360	
gtccgcggcg agggccccgt cgacaaagaa caacgcaccc gcccgcggcg cgcgatctcc	420	
ctgggtcgcg cggtgattgc tgcccttcgc tegccggccg atccgatgt cgggctgtac	480	
5 ggcgccttcg cttacacact ttgttccag ttgcaggatc tgaagcaga ggtgtccccgc	540	
gaacccgacc agcgcgcacat ctgtgtgtac gtccggatc gcctgtggc ctacgatcgc	600	
gcacccggc gcggcgctga catttctac gaattgcgtt ggaaggccca gtccacccgc	660	
ggctgtccgc acgagacccg cgagacgtc tacacccgtt cccgcggca gggtttccgc	720	
gaccacgcgc cgggcgacta tcccaagggtg gtgcgaaaagg cccgcgcggc gtccgcggc	780	
10 ggccgacctgt tcgaggcggtt gcccggccg ctgttcggcg agccatcgca gcccgtggcg	840	
gccgaagtgt tcaagcggtt gtccggatc aaccctgtgc ctatggggg ctgtgtcaat	900	
ctcgcgcacg gcaatttctt ggtgtggcc teccggaaata ttgtcgccg ctccggacgc	960	
cgccggatcg agacctgcgc gatctccgc actatcgccc gcccgtcgta tgcatcgc	1020	
gatgtcgatc agatccagaat gctttgtaaatc tccgagaaagg acgatgtcgatc gctgaatat	1080	
15 tgcacccacg tcgaccgcacaa cgacaaaggcg cgggttgcgtc tgccggcac gatcaaagt	1140	
ctcgcgcgc gccagatcgca gacctattcg aagctgttcc acaccgtcgca tcacgtcgag	1200	
ggcatgtcgac gaccgggtt ccacgcgtc gacgccttcc tcaacccacgc ctgggggttc	1260	
accgtcaccgc ggcgcgcgaa gctgtggggg atgcagttcg tcgaggatca cgacgttagc	1320	
ccggggcgatc ggtatgcggg cgcgttcggc gtgggtcggtc tgatgtgtc gatcaacacc	1380	
20 ggccctaccaa tccgcacccat cggatgtaaagg gacggccgtc cccggatcgatc cgtccggcc	1440	
accctgcgtt tcgacacgcac tccgtcgcc gaggacaaagg aatgcacaggta caaggcccg	1500	
gcactgttcc accggcgctcg cggcgatecc gccaaggccgc tgccgggtt ggcgcggac	1560	
gcacactgttcc cggcaagaaa ggtgtcgctg tgccaccacg acgacatgtt cgtgcacatg	1620	
ctggcgactt atttcaggca ggttgcggcc cggatccacgg tgggtcgctt cgttcaacggc	1680	
25 ctgaagatgt tgccggaaaaa cggatcgatc ttcttggtgc tgccggcccg tcccgcccg	1740	
ccggaggact tcaagatcaa ggatcgatc gacgcgcgc tcgcacaaatc gctggcgatc	1800	
ttccggctttt gcctcgccgtt ccaggcgatc ggcgaatatt ttggcggtac gctcgccag	1860	
ctccgcgcacg cggatcgaccc cggccggctcg cggatccagg tgccggccgg cgcgtgtatc	1920	
cgggtctcc cgaacggatc cccatcgcc cgcgttccatc cgtctatgtt cgcacatgcgc	1980	
30 gacatgcgcg aggacgttgc acgttccgc tccaccgtatc acggatcgatc gatggcgatc	2040	
gagcacaaga ccctgcgggtt cggccggcgatc cggatcgatc gatgtcgatc	2100	
ggccggcgagg tccggcgatc gatgtcgatc aacgccttcgc ggctcgccca ggccggctaa	2160	

<210> 77

35 <211> 733

<212> PRT

<213> Mesorhizobium loti

<400> 77

40 Met Glu Thr Ala Met Thr Met Lys Val Leu Glu Asn Gly Ala Glu Ser

72

1	5	10	15													
Phe	Val	Thr	Ala	Gly	Gly	Ile	Thr	Ile	Thr	Arg	Glu	Arg	His	Asp	Arg	
20																
Pro	Tyr	Ala	Gly	Ala	Ile	Asp	Ala	Tyr	Val	Asp	Gly	Leu	Asn	Ser	Arg	
5	35															
Arg	Gly	Ala	Val	Phe	Ser	Ser	Asn	Tyr	Glu	Tyr	Pro	Gly	Arg	Tyr	Thr	
50																
Arg	Trp	Asp	Thr	Ala	Ile	Ile	Asp	Pro	Pro	Leu	Val	Ile	Ser	Ala	Arg	
65																
10	Gly	Arg	Ala	Met	Arg	Ile	Glu	Ala	Leu	Asn	Arg	Arg	Gly	Glu	Ala	Leu
85																
Leu	Pro	Val	Ile	Gly	Lys	Thr	Leu	Gly	Gly	Leu	Ala	Asp	Ile	Thr	Ile	
100																
Ala	Glu	Thr	Thr	Lys	Thr	Leu	Ile	Arg	Leu	Asp	Val	Ala	Lys	Pro	Gly	
15	115															
Arg	Val	Phe	Thr	Gl	Gl	Gl	Arg	Ser	Arg	Val	Pro	Ser	Val	Phe	Thr	
130																
135																
140																
Val	Leu	Arg	Ala	Ile	Thr	Ala	Leu	Phe	Lys	Thr	Asp	Glu	Asp	Ala	Asn	
145																
150																
155																
160																
20	Leu	Gly	Leu	Tyr	Gly	Ala	Phe	Gly	Tyr	Asp	Leu	Ser	Phe	Gln	Phe	Asp
165																
170																
175																
Pro	Val	Asp	Tyr	Lys	Leu	Glu	Arg	Lys	Pro	Ser	Gln	Arg	Asp	Leu	Val	
180																
185																
190																
Leu	Phe	Leu	Pro	Asp	Glu	Ile	Leu	Val	Val	Asp	His	Tyr	Ser	Ala	Lys	
25	195															
200																
205																
Ala	Trp	Thr	Asp	Arg	Tyr	Asp	Tyr	Ser	Gly	Glu	Gly	Phe	Ser	Thr	Glu	
210																
215																
220																
Gly	Leu	Pro	Arg	Asp	Ala	Ile	Ala	Glu	Pro	Phe	Lys	Thr	Ala	Asp	Arg	
225	.															
230																
235																
240																
30	Ile	Pro	Pro	Arg	Gly	Asp	His	Glu	Pro	Gly	Glu	Tyr	Ala	Asn	Leu	Val
245																
250																
255																
Arg	Arg	Ala	Met	Asp	Ser	Phe	Lys	Arg	Gly	Asp	Leu	Phe	Glu	Val	Val	
260																
265																
270																
Pro	Gly	Gln	Met	Phe	Tyr	Glu	Arg	Cys	Glu	Thr	Gln	Pro	Ser	Asp	Ile	
35	275															
280																
285																
Ser	Arg	Lys	Leu	Lys	Ser	Ile	Asn	Pro	Ser	Pro	Tyr	Ser	Phe	Phe	Ile	
290																
295																
300																
Asn	Leu	Gly	Glu	Asn	Glu	Tyr	Leu	Ile	Gly	Ala	Ser	Pro	Glu	Met	Phe	
305																
310																
315																
320																
40	Val	Arg	Val	Asn	Gly	Arg	Arg	Val	Glu	Thr	Cys	Pro	Ile	Ser	Gly	Thr

73

	325	330	335
Ile Lys Arg Gly Asp Asp Ala Ile Ser Asp Ser Glu Gln Ile Leu Lys			
340	345	350	
Leu Leu Asn Ser Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp			
5 355	360	365	
Val Asp Arg Asn Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Arg			
370	375	380	
Val Ile Gly Arg Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr			
385	390	395	400
10 Val Asp His Ile Glu Gly Arg Leu Arg Glu Gly Met Asp Ala Phe Asp			
405	410	415	
Ala Phe Leu Ser His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys			
420	425	430	
Leu Trp Ala Met Arg Phe Ile Glu Gln Asn Glu Lys Ser Pro Arg Ala			
15 435	440	445	
Trp Tyr Gly Gly Ala Ile Gly Met Val Asn Phe Asn Gly Asp Met Asn			
450	455	460	
Thr Gly Leu Thr Leu Arg Thr Ile Arg Ile Lys Asp Gly Ile Ala Glu			
465	470	475	480
20 Val Arg Ala Gly Ala Thr Leu Leu Phe Asp Ser Ile Pro Glu Glu Glu			
485	490	495	
Glu Ala Glu Thr Glu Leu Lys Ala Ser Ala Met Leu Ser Ala Ile Arg			
500	505	510	
Asp Ala Lys Thr Gly Asn Ser Ala Ser Thr Glu Arg Thr Thr Ala Arg			
25 515	520	525	
Val Gly Asp Gly Val Asn Ile Leu Leu Val Asp His Glu Asp Ser Phe			
530	535	540	
Val His Thr Leu Ala Asn Tyr Phe Arg Gln Thr Gly Ala Asn Val Ser			
545	550	555	560
30 Thr Val Arg Thr Pro Val Pro Asp Glu Val Phe Glu Arg Leu Lys Pro			
565	570	575	
Asp Leu Val Val Leu Ser Pro Gly Pro Gly Thr Pro Lys Asp Phe Asp			
580	585	590	
Cys Ala Ala Thr Ile Arg Arg Ala Arg Ala Arg Asp Leu Pro Ile Phe			
35 595	600	605	
Gly Val Cys Leu Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Glu			
610	615	620	
Leu Arg Gln Leu His Ile Pro Met His Gly Lys Pro Ser Arg Ile Arg			
625	630	635	640
40 Val Ser Lys Pro Gly Ile Ile Phe Ser Gly Leu Pro Lys Glu Val Thr			

74

	645	650	655
Val Gly Arg Tyr His Ser Ile Phe Ala Asp Pro Val Arg Leu Pro Asp			
	660	665	670
Asp Phe Ile Val Thr Ala Glu Thr Glu Asp Gly Ile Ile Met Ala Phe			
5	675	680	685
Glu His Arg Lys Glu Pro Ile Ala Ala Val Gln Phe His Pro Glu Ser			
	690	695	700
Ile Met Thr Leu Gly His Asn Ala Gly Met Arg Ile Ile Glu Asn Ile			
705	710	715	720
10 Val Ala His Leu Pro Arg Lys Ala Lys Glu Lys Ala Ala			
	725	730	

<210> 78			
<211> 732			
15 <212> PRT			
<213> Azospirillum brasiliense			
<400> 78			
Met Tyr Pro Ala Asp Leu Leu Ala Ser Pro Asp Leu Leu Glu Pro Leu			
20 1	5	10	15
Arg Phe Gln Thr Arg Gly Gly Val Thr Val Thr Arg Arg Ala Thr Ala			
	20	25	30
Leu Asp Pro Arg Thr Ala Leu Asp Pro Val Ile Asp Ala Leu Asp Arg			
	35	40	45
25 Arg Arg Gly Leu Leu Leu Ser Ser Gly Val Glu Ala Pro Gly Arg Tyr			
	50	55	60
Arg Arg His Ala Leu Gly Phe Thr Asp Pro Ala Val Ala Leu Thr Ala			
	65	70	75
65 70			80
Arg Gly Arg Thr Leu Arg Ile Asp Ala Leu Asn Gly Arg Gly Gln Val			
	80	85	90
30 85			95
Leu Leu Pro Ala Val Ala Glu Ala Leu Arg Gly Leu Glu Ala Leu Ala			
	100	105	110
Gly Leu Glu Glu Ala Pro Ser Arg Val Thr Ala Ser Ser Ala Ser Pro			
	115	120	125
35 Ala Pro Leu Pro Gly Glu Glu Arg Ser Arg Gln Pro Ser Val Phe Ser			
	130	135	140
Val Leu Arg Ala Val Leu Asp Leu Phe Ala Ala Pro Asp Asp Pro Leu			
	145	150	155
Leu Gly Leu Tyr Gly Ala Phe Ala Tyr Asp Leu Ala Phe Gln Phe Glu			160
	165	170	175

75

Pro Ile Arg Gln Arg Leu Glu Arg Pro Asp Asp Gln Arg Asp Leu Leu
 180 185 190
 Leu Tyr Leu Pro Asp Arg Leu Val Ala Leu Asp Pro Ile Ala Gly Leu
 195 200 205
 5 Ala Arg Leu Val Ala Tyr Glu Phe Ile Thr Ala Ala Gly Ser Thr Glu
 210 215 220
 Gly Leu Glu Cys Gly Gly Arg Asp His Pro Tyr Arg Pro Asp Thr Asn
 225 230 235 240
 Ala Glu Ala Gly Cys Asp His Ala Pro Gly Asp Tyr Gln Arg Val Val
 10 245 250 255
 Glu Ser Ala Lys Ala Ala Phe Arg Arg Gly Asp Leu Phe Glu Val Val
 260 265 270
 Pro Gly Gln Thr Phe Ala Glu Pro Cys Ala Asp Ala Pro Ser Ser Val
 275 280 285
 15 Phe Arg Arg Leu Arg Ala Ala Asn Pro Ala Pro Tyr Glu Ala Phe Val
 290 295 300
 Asn Leu Gly Arg Gly Glu Phe Leu Val Ala Ala Ser Pro Glu Met Tyr
 305 310 315 320
 Val Arg Val Ala Gly Gly Arg Val Glu Thr Cys Pro Ile Ser Gly Thr
 20 325 330 335
 Val Ala Arg Gly Ala Asp Ala Leu Gly Asp Ala Ala Gln Val Leu Arg
 340 345 350
 Leu Leu Thr Ser Ala Lys Asp Ala Ala Glu Leu Thr Met Cys Thr Asp
 355 360 365
 25 Val Asp Arg Asn Asp Lys Ala Arg Val Cys Glu Pro Gly Ser Val Arg
 370 375 380
 Val Ile Gly Arg Arg Met Ile Glu Leu Tyr Ser Arg Leu Ile His Thr
 385 390 395 400
 Val Asp His Val Glu Gly Arg Leu Arg Ser Gly Met Asp Ala Leu Asp
 30 405 410 415
 Ala Phe Leu Thr His Ser Trp Ala Val Thr Val Thr Gly Ala Pro Lys
 420 425 430
 Arg Trp Ala Met Gln Phe Leu Glu Asp Thr Glu Gln Ser Pro Arg Arg
 435 440 445
 35 Trp Tyr Gly Gly Ala Phe Gly Arg Leu Gly Phe Asp Gly Gly Met Asp
 450 455 460
 Thr Gly Leu Thr Leu Arg Thr Ile Arg Met Ala Glu Gly Val Ala Tyr
 465 470 475 480
 Val Arg Ala Gly Ala Thr Leu Leu Ser Asp Ser Asp Pro Asp Ala Glu
 40 485 490 495

76

Asp Ala Glu Cys Arg Leu Lys Ala Ala Ala Phe Arg Asp Ala Ile Arg
500 505 510
Gly Thr Ala Ala Gly Ala Ala Pro Thr Leu Pro Ala Ala Pro Arg Gly
515 520 525
5 Gly Glu Gly Arg Arg Val Leu Leu Val Asp His Asp Asp Ser Phe Val
530 535 540
His Thr Leu Ala Asp Tyr Leu Arg Gln Thr Gly Ala Ser Val Thr Thr
545 550 555 560
Leu Arg His Ser His Ala Arg Ala Ala Leu Ala Glu Arg Arg Pro Asp
10 565 570 575
Leu Val Val Leu Ser Pro Gly Pro Gly Arg Pro Ala Asp Phe Asp Val
580 585 590
Ala Gly Thr Ile Asp Ala Ala Leu Ala Leu Gly Leu Pro Val Phe Gly
595 600 605
15 Val Cys Leu Gly Leu Gln Gly Met Val Glu Arg Phe Gly Gly Ala Leu
610 615 620
Asp Val Leu Pro Glu Pro Val His Gly Lys Ala Thr Glu Val Arg Val
625 630 635 640
Leu Gly Gly Ala Leu Phe Ala Gly Leu Pro Glu Arg Leu Thr Val Gly
20 645 650 655
Arg Tyr His Ser Leu Val Ala Arg Arg Asp Arg Leu Pro Ala Asp Leu
660 665 670
Thr Val Thr Ala Glu Thr Ala Asp Gly Leu Val Met Ala Val Glu His
675 680 685
25 Arg Arg Leu Pro Leu Ala Ala Val Gln Phe His Pro Glu Ser Ile Leu
690 695 700
Ser Leu Asp Gly Gly Ala Gly Leu Ala Leu Leu Gly Asn Val Met Asp
705 710 715 720
Arg Leu Ala Ala Gly Ala Leu Thr Asp Ala Ala Ala
30 725 730

<210> 79
<211> 731
<212> PRT
35 <213> Brucella melitensis

<400> 79
Met Asn Ala Lys Thr Ala Asp Ser Glu Ile Phe Gln His Glu Thr Ala
1 5 10 15
40 Gly Gly Ile Ile Val Glu Arg Val Arg His Leu Thr Ala Tyr Lys Gly

77

	20	25	30
Ala Ile Glu Ser Tyr Ile Asp Val Leu Asn Glu Trp Arg Gly Ala Val			
	35	40	45
Phe Ser Ser Asn Tyr Glu Tyr Pro Gly Arg Tyr Thr Arg Trp Asp Thr			
	50	55	60
Ala Ile Val Asp Pro Pro Val Val Ile Thr Ser Arg Ala Arg Thr Met			
	65	70	75
Arg Ile Glu Ala Leu Asn Ala Arg Gly Val Ile Leu Leu Arg Pro Ile			
	85	90	95
10 Leu Asp Thr Val Lys Ala Leu Ser Glu Val Lys Ile Asp Gln Ser Gly			
	100	105	110
Glu Asn Arg Ile Asp Leu Thr Ile Val Glu Pro Val Gly Thr Phe Thr			
	115	120	125
Glu Glu Glu Arg Ser Arg Met Pro Ser Val Phe Thr Val Leu Arg Ala			
	130	135	140
Ile Val Gly Leu Phe Phe Ser Glu Glu Asp Ala Asn Leu Gly Leu Tyr			
	145	150	155
Gly Ala Phe Gly Tyr Asp Leu Ala Phe Gln Phe Asp Pro Ile Gln Tyr			
	165	170	175
20 Lys Leu Lys Arg Pro Asp Asp Gln Arg Asp Leu Val Leu Phe Ile Pro			
	180	185	190
Asp Glu Ile Phe Val Ala Asp His Tyr Ala Ala Arg Ala Trp Val Asp			
	195	200	205
Arg Tyr Glu Phe Arg Cys Gly Ser Ser Thr His Gly Leu Asp Arg			
	210	215	220
Ala Thr Pro Val Val Pro Phe Lys Pro Ser Glu Arg Lys Leu Ala Arg			
	225	230	235
Gly Asp His Asn Pro Gly Glu Tyr Ala Arg Leu Val Glu Arg Ala Lys			
	245	250	255
30 Glu Ser Phe Lys Arg Gly Asp Leu Phe Glu Val Val Pro Gly Gln Thr			
	260	265	270
Phe Tyr Glu Arg Cys His Thr Ala Pro Ser Glu Ile Phe Arg Arg Leu			
	275	280	285
Lys Ser Ile Asn Pro Ser Pro Tyr Ser Phe Phe Ile Asn Leu Gly Glu			
	290	295	300
Ser Glu Tyr Leu Val Gly Ala Ser Pro Glu Met Phe Val Arg Val Asn			
	305	310	315
Gly Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Lys Arg Gly			
	325	330	335
40 Glu Asp Ala Ile Ser Asp Ser Glu Gln Ile Leu Lys Leu Leu Asn Ser			

78

340	345	350
Lys Lys Asp Glu Ser Glu Leu Thr Met Cys Ser Asp Val Asp Arg Asn		
355	360	365
Asp Lys Ser Arg Val Cys Glu Pro Gly Ser Val Arg Val Ile Gly Arg		
5 370	375	380
Arg Gln Ile Glu Met Tyr Ser Arg Leu Ile His Thr Val Asp His Ile		
385	390	395
Glu Gly Arg Leu Arg Asp Gly Met Asp Ala Phe Asp Gly Phe Leu Ser		
405	410	415
10 His Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met		
420	425	430
Arg Phe Leu Glu Glu Asn Glu Arg Ser Pro Arg Ala Trp Tyr Gly Gly		
435	440	445
Ala Ile Gly Met Met His Phe Asn Gly Asp Met Asn Thr Gly Leu Thr		
15 450	455	460
Leu Arg Thr Ile Arg Ile Lys Asp Gly Val Ala Glu Ile Arg Ala Gly		
465	470	475
Ala Thr Leu Leu Phe Asp Ser Asn Pro Asp Glu Glu Ala Glu Thr		
485	490	495
20 Glu Leu Lys Ala Ser Ala Met Ile Ala Ala Val Arg Asp Ala Gln Lys		
500	505	510
Ser Asn Gln Ile Ala Glu Glu Ser Val Ala Ala Lys Val Gly Glu Gly		
515	520	525
Val Ser Ile Leu Leu Val Asp His Glu Asp Ser Phe Val His Thr Leu		
25 530	535	540
Ala Asn Tyr Phe Arg Gln Thr Gly Ala Lys Val Ser Thr Val Arg Ser		
545	550	555
Pro Val Ala Glu Glu Ile Phe Asp Arg Val Asn Pro Asp Leu Val Val		
565	570	575
30 Leu Ser Pro Gly Pro Gly Ser Pro Gln Asp Phe Asp Cys Lys Ala Thr		
580	585	590
Ile Asp Lys Ala Arg Lys Arg Gln Leu Pro Ile Phe Gly Val Cys Leu		
595	600	605
Gly Leu Gln Ala Leu Ala Glu Ala Tyr Gly Gly Ala Leu Arg Gln Leu		
35 610	615	620
Arg Val Pro Val His Gly Lys Pro Ser Arg Ile Arg Val Ser Lys Pro		
625	630	635
Glu Arg Ile Phe Ser Gly Leu Pro Glu Glu Val Thr Val Gly Arg Tyr		
645	650	655
40 His Ser Ile Phe Ala Asp Pro Glu Arg Leu Pro Asp Asp Phe Leu Val		

79

660	665	670
Thr Ala Glu Thr Glu Asp Gly Ile Ile Met Ala Phe Glu His Lys His		
675	680	685
Glu Pro Val Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr Leu		
5 690	695	700
Gly His Asn Ala Gly Met Arg Met Ile Glu Asn Ile Val Thr His Leu		
705	710	715
Ala Gly Lys His Lys Ala Arg Arg Thr Asn Tyr		
725	730	
 10		
<210> 80		
<211> 735		
<212> PRT		
<213> Nostoc sp.		
 15		
<400> 80		
Met Ile Ala Asp Ser His Ser Tyr Arg Thr Asn Gly Asn Val Arg Val		
1	5	10
Ser Arg Ser Ile Thr Gln Val Lys Met Glu Thr Ala Leu Glu Glu Ile		
20 20	25	30
Leu Phe Tyr Leu Asn Ser Gln Arg Gly Gly Leu Leu Thr Ser Ser Tyr		
35	40	45
Glu Tyr Pro Gly Arg Tyr Lys Arg Trp Ala Ile Gly Phe Val Asn Pro		
50	55	60
25 Pro Val Glu Leu Ser Thr Ser Gly Asn Thr Phe Thr Leu Thr Ala Leu		
65	70	75
Asn Glu Arg Gly Tyr Val Leu Leu Pro Val Ile Phe Glu Cys Leu Ser		
85	90	95
Lys Ser Glu Gln Leu Gln Lys Leu Thr Glu His His His Lys Ile Thr		
30 100	105	110
Gly Leu Val Lys Ser Thr Pro Glu Phe Phe Ala Glu Glu Glu Arg Ser		
115	120	125
Lys Gln Pro Ser Thr Phe Thr Val Ile Arg Glu Ile Leu His Ile Phe		
130	135	140
35 Ser Ser Gln Glu Asp Glu His Leu Gly Leu Tyr Gly Ala Phe Gly Tyr		
145	150	155
Asp Leu Val Phe Gln Phe Glu Gln Ile Thr Gln Cys Leu Glu Arg Pro		
165	170	175
Gln Asp Gln Arg Asp Leu Val Leu Tyr Leu Pro Asp Glu Leu Ile Val		
40 180	185	190

80

Val Asp Tyr Tyr Gln Gln Gln Ala Phe Arg Leu Glu Tyr Asp Phe Ile
 195 200 205
 Thr Ala His Gly Ser Thr Tyr Asp Leu Pro Arg Thr Gly Glu Ser Val
 210 215 220
 5 Asp Tyr Arg Gly Gln Cys Leu Thr Pro Pro Gln Asn Ala Asp His Lys
 225 230 235 240
 Ile Gly Glu Tyr Ala Lys Leu Val Glu Phe Ala Leu Asp Tyr Phe Arg
 245 250 255
 Arg Gly Asp Leu Phe Glu Val Val Pro Ser Gln Asn Phe Phe Thr Ala
 10 260 265 270
 Cys Glu Ala Pro Pro Ser Gln Leu Phe Glu Thr Leu Lys Gln Ile Asn
 275 280 285
 Pro Ser Pro Tyr Gly Phe Ile Phe Asn Leu Gly Gly Glu Tyr Ile Ile
 290 295 300
 15 Gly Ala Ser Pro Glu Met Phe Val Arg Val Glu Gly Arg Arg Val Glu
 305 310 315 320
 Thr Cys Pro Ile Ser Gly Thr Ile Thr Arg Gly His Asp Ala Ile Asp
 325 330 335
 Asp Ala Val Gln Ile Arg Gln Leu Leu Asn Ser His Lys Asp Glu Ala
 20 340 345 350
 Glu Leu Thr Met Cys Thr Asp Val Asp Arg Asn Asp Lys Ser Arg Ile
 355 360 365
 Cys Glu Pro Gly Ser Val Lys Val Ile Gly Arg Arg Gln Ile Glu Leu
 370 375 380
 25 Tyr Ser His Leu Ile His Thr Val Asp His Val Glu Gly Ile Leu Arg
 385 390 395 400
 Pro Glu Phe Asp Ala Leu Asp Ala Phe Leu Ser His Thr Trp Ala Val
 405 410 415
 Thr Val Thr Gly Ala Pro Lys Arg Ala Ala Ile Gln Phe Ile Glu Lys
 30 420 425 430
 Asn Glu Arg Ser Val Arg Arg Trp Tyr Gly Gly Ala Val Gly Tyr Leu
 435 440 445
 Asn Phe Asn Gly Asn Leu Asn Thr Gly Leu Ile Leu Arg Thr Ile Arg
 450 455 460
 35 Leu Gln Asp Ser Ile Ala Glu Val Arg Val Gly Ala Thr Leu Leu Tyr
 465 470 475 480
 Asp Ser Ile Pro Gln Ala Glu Glu Gln Glu Thr Ile Thr Lys Ala Ala
 485 490 495
 Ala Ala Phe Glu Thr Ile Arg Arg Ala Lys Gln Ile Asp Pro Gln Ile
 40 500 505 510

81

Glu Glu Ser Ser Thr Arg Lys Leu Ser Lys Tyr Leu Pro Asp Gly Gln
 515 520 525
 Ser Gly Lys His Ile Leu Leu Ile Asp His Glu Asp Ser Phe Val His
 530 535 540
 5 Thr Leu Ala Asn Tyr Ile Arg Ser Thr Gly Ala Thr Val Thr Thr Leu
 545 550 555 560
 Arg His Gly Phe Ser Glu Ser Leu Phe Asp Thr Glu Arg Pro Asp Leu
 565 570 575
 Val Val Leu Ser Pro Gly Pro Gly Arg Pro Ser Glu Phe Lys Val Gln
 10 580 585 590
 Glu Thr Val Ala Ala Cys Val Arg Arg Gln Ile Pro Leu Phe Gly Val
 595 600 605
 Cys Leu Gly Leu Gln Gly Ile Val Glu Ala Phe Gly Gly Glu Leu Gly
 610 615 620
 15 Val Leu Asn Tyr Pro Gln His Gly Lys Ser Ser Arg Ile Phe Val Thr
 625 630 635 640
 Ala Pro Asp Ser Val Met Phe Gln Asp Leu Pro Glu Ser Phe Thr Val
 645 650 655
 Gly Arg Tyr His Ser Leu Phe Ala Leu Ser Gln Arg Leu Pro Lys Glu
 20 660 665 670
 Leu Lys Val Thr Ala Ile Ser Asp Asp Glu Val Ile Met Ala Ile Glu
 675 680 685
 His Gln Thr Leu Pro Ile Ala Ala Val Gln Phe His Pro Glu Ser Ile
 690 695 700
 25 Met Thr Leu Ala Gly Glu Val Gly Leu Met Met Ile Lys Asn Val Val
 705 710 715 720
 Gln Lys Tyr Thr Gln Ser Gln Gln Ser Thr Val Pro Ile Tyr Asp
 725 730 735

 30 <210> 81
 <211> 715
 <212> PRT
 <213> Nostoc sp.

 35 <400> 81
 Met Arg Val Ser Arg Ser Thr Thr Glu Val Lys Met Asp Thr Ala Leu
 1 5 10 15
 Asp Glu Ile Leu Phe His Leu Asn Gln Val Arg Gly Gly Leu Leu Thr
 20 25 30
 40 Ser Ser Tyr Glu Tyr Pro Gly Arg Tyr Lys Arg Trp Ala Ile Gly Phe

82

35	40	45
Ile Asn Pro Pro Leu Gln Leu Thr Thr Arg Glu Asn Ala Phe Thr Ile		
50	55	60
Ser Ser Leu Asn Pro Arg Gly Gln Val Leu Leu Pro Thr Leu Phe Gln		
55	70	75
His Leu Ser Ala Gln Ser Gln Leu Gln Ile Ser Leu Asn His Asp		80
85	90	95
Tyr Ile Thr Gly Glu Ile Arg Pro Thr Lys Gln Leu Phe Thr Glu Glu		
100	105	110
10 Gln Arg Ser Lys Gln Pro Ser Ala Phe Thr Val Ile Arg Glu Ile Leu		
115	120	125
Gln Ile Phe Ala Ser Asp Glu Asp Glu His Leu Gly Leu Tyr Gly Ala		
130	135	140
Phe Gly Tyr Asp Leu Val Phe Gln Phe Glu Pro Ile Pro Gln Lys Ile		
15 145	150	155
Ala Arg Pro Ala Asp Gln Arg Asp Leu Val Leu Tyr Leu Pro Asp Glu		160
165	170	175
Leu Ile Val Val Asp Tyr Tyr Leu Gln Lys Ala Tyr Arg His Gln Tyr		
180	185	190
20 Glu Phe Ala Thr Glu His Gly Asn Thr Glu His Leu Pro Arg Thr Gly		
195	200	205
Gln Ser Ile Asp Tyr Gln Gly Lys His Leu Leu Pro Asn Gln Thr Ala		
210	215	220
Asp His Gln Pro Gly Glu Tyr Ala Asn Leu Val Glu Gln Ala Leu Asp		
25 225	230	235
Tyr Phe Arg Arg Gly Asp Leu Phe Glu Val Val Pro Ser Gln Asn Phe		240
245	250	255
Phe Thr Ala Cys Glu Gln Ser Pro Ser Gln Leu Phe Gln Thr Leu Arg		
260	265	270
30 Gln Ile Asn Pro Ser Pro Tyr Gly Phe Leu Leu Asn Leu Gly Glu		
275	280	285
Tyr Leu Ile Gly Ala Ser Pro Glu Met Phe Val Arg Val Asp Gly Arg		
290	295	300
Arg Val Glu Thr Cys Pro Ile Ser Gly Thr Ile Arg Arg Gly Glu Asp		
35 305	310	315
Ala Leu Gly Asp Ala Val Gln Ile Arg Gln Leu Leu Asn Ser His Lys		320
325	330	335
Asp Glu Ala Glu Leu Thr Met Cys Thr Asp Val Asp Arg Asn Asp Lys		
340	345	350
40 Ser Arg Ile Cys Glu Pro Gly Ser Val Arg Val Ile Gly Arg Arg Gln		

83

355	360	365
Ile Glu Leu Tyr Ser His Leu Ile His Thr Val Asp His Val Glu Gly		
370	375	380
Ile Leu Arg Pro Glu Phe Asp Ala Leu Asp Ala Phe Leu Ser His Thr		
5 385	390	395
Trp Ala Val Thr Val Thr Gly Ala Pro Lys Arg Ala Ala Met Gln Phe		400
405	410	415
Ile Glu Gln His Glu Arg Ser Ala Arg Arg Trp Tyr Gly Gly Ala Val		
420	425	430
10 Gly Tyr Leu Gly Phe Asn Gly Asn Leu Asn Thr Gly Leu Thr Leu Arg		
435	440	445
Thr Ile Arg Leu Gln Asp Ser Ile Ala Glu Val Arg Val Gly Ala Thr		
450	455	460
Val Leu Tyr Asp Ser Ile Pro Ser Ala Glu Glu Glu Glu Thr Ile Thr		
15 465	470	475
Lys Ala Thr Ala Leu Phe Glu Thr Ile Arg Arg His Thr Thr Ala Asn		480
485	490	495
Lys Thr Gln Gly Asn Asp Ser His Arg Pro Gly Asp Ile Ala His Asn		
500	505	510
20 Lys Arg Ile Leu Leu Ile Asp Tyr Glu Asp Ser Phe Val His Thr Leu		
515	520	525
Ala Asn Tyr Ile Arg Thr Thr Gly Ala Thr Val Thr Thr Leu Arg His		
530	535	540
Gly Phe Ala Glu Ser Tyr Phe Asp Ala Glu Arg Pro Asp Leu Val Val		
25 545	550	555
Leu Ser Pro Gly Pro Gly Arg Pro Ser Asp Phe Arg Val Pro Gln Thr		560
565	570	575
Val Ala Ala Leu Val Gly Arg Glu Ile Pro Ile Phe Gly Val Cys Leu		
580	585	590
30 Gly Leu Gln Gly Ile Val Glu Ala Phe Gly Gly Glu Leu Gly Val Leu		
595	600	605
Asp Tyr Pro Gln His Gly Lys Pro Ala Arg Ile Ser Val Thr Ala Pro		
610	615	620
Asp Ser Val Leu Phe Gln Asn Leu Pro Ala Ser Phe Ile Val Gly Arg		
35 625	630	635
Tyr His Ser Leu Phe Ala Gln Pro Gln Thr Ile Pro Gly Glu Leu Lys		640
645	650	655
Val Thr Ala Ile Ser Glu Asp Asn Val Ile Met Ala Ile Glu His Gln		
660	665	670
40 Thr Leu Pro Ile Ala Ala Val Gln Phe His Pro Glu Ser Ile Met Thr		

84

675	680	685
Leu Ala Gly Glu Val Gly Gln Thr Ile Ile Lys Asn Val Val Gln Thr		
690	695	700
Tyr Thr Gln Thr Leu Glu Thr Ser Ile Tyr Ser		
5 705	710	715
<210> 82		
<211> 719		
<212> PRT		
10 <213> Rhodopseudomonas palustris		
<400> 82		
Met Asn Arg Thr Val Phe Ser Leu Pro Ala Thr Ser Asp Tyr Lys Thr		
1	5	10
15 Ala Ala Gly Leu Ala Val Thr Arg Ser Ala Gln Pro Phe Ala Gly Gly		
20	25	30
Gln Ala Leu Asp Glu Leu Ile Asp Leu Leu Asp His Arg Arg Gly Val		
35	40	45
Met Leu Ser Ser Gly Thr Thr Val Pro Gly Arg Tyr Glu Ser Phe Asp		
20	50	55
Leu Gly Phe Ala Asp Pro Pro Leu Ala Leu Thr Thr Arg Ala Glu Lys		
65	70	75
Phe Thr Ile Glu Ala Leu Asn Pro Arg Gly Arg Val Leu Ile Ala Phe		
85	90	95
25 Leu Ser Asp Lys Leu Glu Glu Pro Cys Val Val Val Glu Gln Ala Cys		
100	105	110
Ala Thr Lys Ile Arg Gly His Ile Val Arg Gly Glu Ala Pro Val Asp		
115	120	125
Glut Glu Gln Arg Thr Arg Arg Ala Ser Ala Ile Ser Leu Val Arg Ala		
30	130	135
Val Ile Ala Ala Phe Ala Ser Pro Ala Asp Pro Met Leu Gly Leu Tyr		
145	150	155
Gly Ala Phe Ala Tyr Asp Leu Val Phe Gln Phe Glu Asp Leu Lys Gln		
165	170	175
35 Lys Arg Ala Arg Glu Ala Asp Gln Arg Asp Ile Val Leu Tyr Val Pro		
180	185	190
Asp Arg Leu Leu Ala Tyr Asp Arg Ala Thr Gly Arg Gly Val Asp Ile		
195	200	205
Ser Tyr Glu Phe Ala Trp Lys Gly Gln Ser Thr Ala Gly Leu Pro Asn		
40	210	215
		220

85

Glu Thr Ala Glu Ser Val Tyr Thr Gln Thr Gly Arg Gln Gly Phe Ala
 225 230 235 240
 Asp His Ala Pro Gly Asp Tyr Pro Lys Val Val Glu Lys Ala Arg Ala
 245 250 255
 5 Ala Phe Ala Arg Gly Asp Leu Phe Glu Ala Val Pro Gly Gln Leu Phe
 260 265 270
 Gly Glu Pro Cys Glu Arg Ser Pro Ala Glu Val Phe Lys Arg Leu Cys
 275 280 285
 Arg Ile Asn Pro Ser Pro Tyr Gly Gly Leu Leu Asn Leu Gly Asp Gly
 10 290 295 300
 Glu Phe Leu Val Ser Ala Ser Pro Glu Met Phe Val Arg Ser Asp Gly
 305 310 315 320
 Arg Arg Ile Glu Thr Cys Pro Ile Ser Gly Thr Ile Ala Arg Gly Val
 325 330 335
 15 Asp Ala Ile Ser Asp Ala Glu Gln Ile Gln Lys Leu Leu Asn Ser Glu
 340 345 350
 Lys Asp Glu Phe Glu Leu Asn Met Cys Thr Asp Val Asp Arg Asn Asp
 355 360 365
 Lys Ala Arg Val Cys Val Pro Gly Thr Ile Lys Val Leu Ala Arg Arg
 20 370 375 380
 Gln Ile Glu Thr Tyr Ser Lys Leu Phe His Thr Val Asp His Val Glu
 385 390 395 400
 Gly Met Leu Arg Pro Gly Phe Asp Ala Leu Asp Ala Phe Leu Thr His
 405 410 415
 25 Ala Trp Ala Val Thr Val Thr Gly Ala Pro Lys Leu Trp Ala Met Gln
 420 425 430
 Phe Val Glu Asp His Glu Arg Ser Pro Arg Arg Trp Tyr Ala Gly Ala
 435 440 445
 Phe Gly Val Val Gly Phe Asp Gly Ser Ile Asn Thr Gly Leu Thr Ile
 30 450 455 460
 Arg Thr Ile Arg Met Lys Asp Gly Leu Ala Glu Val Arg Val Gly Ala
 465 470 475 480
 Thr Cys Leu Phe Asp Ser Asn Pro Val Ala Glu Asp Lys Glu Cys Gln
 485 490 495
 35 Val Lys Ala Ala Ala Leu Phe Gln Ala Leu Arg Gly Asp Pro Ala Lys
 500 505 510
 Pro Leu Ser Ala Val Ala Pro Asp Ala Thr Gly Ser Gly Lys Lys Val
 515 520 525
 Leu Leu Val Asp His Asp Asp Ser Phe Val His Met Leu Ala Asp Tyr
 40 530 535 540

86

Phe Arg Gln Val Gly Ala Gln Val Thr Val Val Arg Tyr Val His Gly
 545 550 555 560
 Leu Lys Met Leu Ala Glu Asn Ser Tyr Asp Leu Leu Val Leu Ser Pro
 565 570 575
 5 Gly Pro Gly Arg Pro Glu Asp Phe Lys Ile Lys Asp Thr Ile Asp Ala
 580 585 590
 Ala Leu Ala Lys Lys Leu Pro Ile Phe Gly Val Cys Leu Gly Val Gln
 595 600 605
 Ala Met Gly Glu Tyr Phe Gly Gly Thr Leu Gly Gln Leu Ala Gln Pro
 10 610 615 620
 Ala His Gly Arg Pro Ser Arg Ile Gln Val Arg Gly Gly Ala Leu Met
 625 630 635 640
 Arg Gly Leu Pro Asn Glu Val Thr Ile Gly Arg Tyr His Ser Leu Tyr
 645 650 655
 15 Val Asp Met Arg Asp Met Pro Lys Glu Leu Thr Val Thr Ala Ser Thr
 660 665 670
 Asp Asp Gly Ile Ala Met Ala Ile Glu His Lys Thr Leu Pro Val Gly
 675 680 685
 Gly Val Gln Phe His Pro Glu Ser Leu Met Ser Leu Gly Gly Glu Val
 20 690 695 700
 Gly Leu Arg Ile Val Glu Asn Ala Phe Arg Leu Gly Gln Ala Ala
 705 710 715

<210> 83
 25 <211> 2160
 <212> DNA
 <213> Rhodopseudomonas palustris

<400> 83
 30 atgaacagaga ccgttttctc gettcccgcg accagcgact ataagaccgc cgccggcctc 60
 gccgtgacgc gcacgcgcgc gcctttgcc ggccggccagg cgctcgacgc gctgtatcgat 120
 ctgctcgacc accgcgcgggg cgtgtatcgct tcgtccggca caaccgtgc gggccgcatac 180
 gagagcttcg acctcggtt tcccgatccg ccgcgtggcgc tcaccactag ggccgaaaaaa 240
 ttacccatcg aggcgcetcaa tccgcgcggc cgggtgtcga tcgcgttctt gtccgacaag 300
 35 cttgaagagc cctcgctgtt ggtggaggcag gcctgcgcga ccaagatcg gggccacatc 360
 gtccgcgcgcg aggccccggc cgacgaagaa caacgcaccc gccgcgcgcgc cgcgatctcc 420
 ctgggtgcgcg cgggtattgc tgcccttcgc tcggccggcc atcccgatgtc cgggtgtac 480
 ggccgcctcg cctacgaccc tgggttccag ttcgaggatc tgaaggcaga ggcgtccgcgc 540
 gaagccgcacc agcgcgcacat cgtgtgtac gtgcggatc gcctgtggc ctacgatcg 600
 40 gccacccggc gccggcgatcg catttcctac gaattcgctt ggaaggccca gtccacccgc 660

87

ggcctgcca acgagaccgc	cgagagcgtc tacaccaga	ccggccggca gggtttcgcc	720
gaccacccc	cgggcgacta tccaaagggt	gtcgagaagg cccgcgcggc	780
ggcacatgt	tcgaggcggt	gccccggcag ctgttccggc	840
gccagaatgt	tcaagcggtt	agccatgcga ggggtcgccg	900
5 ctcggcagc	gcaattcc	gtgtcgccc tgcggaaaa	960
cgccggatcg	acacctgccc	tgttgcgtcc ctcggacggc	1020
gatgtcgacg	gatccaga	gatctccggc actatgcgc	1080
tgacccgacg	tcgaccgcaa	cgacaaaggcg	1140
ctcgccgcgc	gccagatcga	cgggtcgcc tgccggcac	1200
10 gcatgtcg	gaccgggtt	gacccgttcc teacccae	1260
acgttcaccc	cgccggcgaa	ctgggtggcg atgcagttcg	1320
ccggcgccgt	gttatgcggg	tcgaggatca cgagctgatc	1380
ggcctacca	ttccgcacca	ccggatgaa	1440
acctgcctt	tcggcgcc	gacggcctcg	1500
15 gactgttcc	aggcgctcg	ccggcgttcc	1560
gcaactgg	cgccatccc	tgccggcggt	1620
ctggcgact	atttcaggca	ggtgcgtcg	1680
ctgaagatgc	ttggcgaaaa	gtcgaccac	1740
ccggaggact	tcaagatcaa	cgatgcgtc	1800
20 ttccggcgat	ccaggcgatg	ggcgaatatt	1860
ctcgccgc	cggtca	ttggcggtac	1920
cgcgttcc	cgacagggt	tgccggcg	1980
gacatcgca	aggagctgac	cgatctatgt	2040
gagcacaaga	ccctggcggt	cgatccacc	2100
25 ggccgcgagg	tcggcgctcg	cccgatgtcg	2160

<210> 84

<211> 2190

<212> DNA

30 <213> Artificial Sequence

<220>

<223> An A. tumefaciens mutant.

35 <400> 84

atggtaacga	tcattcagga	tgacggagcg	gagacctacg	agacgaaagg	cgccatccag	60
gtcagccgaa	agcgccggcc	caccgattt	gccaacgcca	tcgataatta	catcgaaaag	120
cttgattccc	atcgccggcgc	gttttttcg	tccaaactatg	aatatccggg	cggttacacc	180
cgttggata	cgccatcgat	cgatccgccc	ctggccat	cctgttttgg	ccgcaagatg	240
40 tggatcgaa	cctataatgg	ccggccggaa	gtgctgtcg	atttcattac	ggaaaagctt	300

aaggcgacac	ccgatctcac	cctcgccgt	tcctcgaccc	gcggcgctcg	tcttaccgtc	360	
aacgaacccg	accgtgttt	caccgaagaa	gaacgtcg	aaatcccgc	ggtgttccac	420	
gctctcagag	ccatcgctcg	cctttctat	tcgagcgccg	attcggccat	cggccgttgc	480	
ggtgccctcg	gttacgtatc	cgccttccag	ttcgacgcga	tcaagcttc	gctggcgcgt	540	
5	ccggaaagacc	agcggtgacat	ggtgctgttt	ctgcccgtg	aaatccctgt	cgttgatcac	600
tatccgcga	aggcctggat	cgaccgttac	gattcgaga	aggacggcat	gacgacggac	660	
ggaaaatctt	ccgacattac	ccccgatccc	ttaaagacca	ccgataccat	cccgcccaag	720	
ggcgatcacc	gtccggcg	atattccgag	cttggtgtga	aggccaaggaa	aagtttccgc	780	
cgcggegacc	tgttgcggat	cgttcccgcc	cagaattatca	ttgagcgttg	cgaaaagcaat	840	
10	ccgtcggega	tttcccgcc	ctctgaaggcg	atcaaccatc	cccccatttc	cttcttcatc	900
aattcgtgg	atccgagaata	tcgttgcgg	gcctcggcg	aatgtttcg	gcgcgtctcc	960	
ggccgtcgca	tcgagacccgt	cccgatataca	ggcaccatca	agcgccggcg	cgatccgatt	1020	
gccgacagcg	agcagatttt	gaaactgttc	aactcgaaaa	aggacgaaatc	cgaactgacc	1080	
atgtgtctcg	acgttgacccg	caacgacaag	agccggctt	gcgagccggg	ttcggtaag	1140	
15	gtcattggcc	gccgcacat	cgagatgtat	tcacgcctca	tccacaccgt	cgatcacatc	1200
gaaggccgccc	tgccgcacga	tatggacgcc	tttgacgggt	tcctcagcca	ccgcctggcc	1260	
gtcaccgtca	ccgtgtcacc	aaagctgtgg	gcatgcgtc	tcatcgaaagg	tcatgaaaag	1320	
accccgccg	cctggatattg	cggtgcgatc	ggcatggtcg	gttcaacagg	cgacatgaat	1380	
accggcctga	cgctgcgcac	catccggatc	aaggacggtt	ttgcgaaagt	gcgcgcggc	1440	
20	gcgaccctgc	tcaatgttac	caaccccgag	gaagaagaa	ccgaaacccg	actgaaggcc	1500
tccgcctatga	tatcagccat	tcgtgacgc	aaaggcacca	acttgtccgc	caccaagggt	1560	
gatggcccca	aagtgcggcac	cggcgtcaca	atccgtctgc	tcgaccacga	agacagcttc	1620	
gtgcacacgc	ttggcgaata	tttccggcc	acggggcgca	eggtctcgac	cgtcagatca	1680	
ccggtcgc	ccgacgtgtt	cgatcgatc	cageccggacc	tcgttgcet	gtcgcggga	1740	
25	ccggcagace	cgacgtgattt	cgactcgaa	tcgaaacgtca	aggccggccg	cgcccgcgat	1800
ctgcccgtat	tcggcggtt	cctcggtcg	caggcatttg	cagaaggcta	ttggccgcag	1860	
ctgcccggc	ttgtgtgtcc	catgcacggc	aaggcttcgc	gcaccccggt	gttggaaacc	1920	
ggccctgtt	tctccggctt	cgcaaggaa	gtcacggctg	tcgttacca	ttcgatctc	1980	
gcgcgtcccg	ccacccctgc	gggtgatttc	atcatcacc	cagaaggcg	ggacggcacc	2040	
30	atcatggc	tcgaacacgc	caaggaaaccg	ttgcggccgg	ttcagttcca	cccgaaatcg	2100
atcatgacgc	tcggacagga	cgcggccatg	cggatgtatcg	agaatgtcg	gttgcacatcg	2160	
acccgcaagg	cgaagaccaa	ggccgcgtg				2190	

<210> 85

35 <211> 2190

<212> DNA

<213> Artificial Sequence

<220>

40 <223> An A. tumefaciens mutant.

<400> 85

atggtaacga tcattcagga tgacggagcg gagacctacg agacgaaagg cggcatccag	60
gtcagccgaa agcgcggcc caccgattat gccaacgcga tcgataatta catcgaaaag	120
5 ctgttcccc atcgccggcgt gtattttcg tccaaactatg aatatccggg cggttacacc	180
cgtctggata cgcccatcgat cgatccggcg ctggcattt cctgttttg ccgcaagtg	240
tggatcgaa cctataatgg ccgcggcgaat tgctgtctg atttcattac ggaaaaagtc	300
aaggcgacac ccgatctcaat cttccggcgt ctctcgaccg gccggctcgat ttttacccgtc	360
aacaacccgg accgtgtctt caccgaagaa gaacgtcgaa aaatcccgac ggttccacc	420
10 gtctcagac ccatecgatc ctcttttatc tcgagccggg attccggccat cggcttgc	480
ggtcgttccgtt gtatcgatctt cgccttcccgat tcgacgcgtca tcaagcttgcgt	540
cggaaagacc acggctgtatc gtgtgtgtttt ctggccgtatc aaatctcgatc cttttatcc	600
tattccggca aggccgttatc cgaccgttac gatttcgaga aggacggcat gacgcggac	660
ggaaaatctt ccgcattaccc cccgatccc ttcaagacca ccgtatccat cccggccaa	720
15 ggccatcacc gtccggcga atattccgag ctgttgttgcg aggccaaggaa aagcttccgc	780
cggccggaccat tttccggcgtt ctttccggc cagaattatca tggagcgttgcgaaat	840
cggccggcgtt tttccggcgtt ctttccggc ctttccggc ctttccggc ctttccggc	900
aatctcgccgtt atcggaaata tctgttgcgc tcttcggccgcgaaatgttgcgttgcgttgc	960
ggccgtcgca tcgagacccatc cccgatataca ggcaccatca agcgcggcga cgatcccgat	1020
20 ggcacacggc agcagatttt gaaactgttc aactcgaaaaggacgaaatc cgaactggacc	1080
atgtgtcggtt acgtggaccc caacgacaaagccggatcgatccatc tttccggatcgatccatc	1140
gttattggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1200
gaaggccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1260
gtcaccgttca cccgatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1320
25 accggccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1380
accggccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1440
gcccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1500
tccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1560
gatggccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1620
30 gtgcacacgc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1680
ccggccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1740
ccggccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1800
ctggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1860
ctggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1920
35 ggcctcgatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	1980
gcccggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	2040
atcatggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	2100
atcatggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	2160
accggccatccatc tttccggatcgatccatc tttccggatcgatccatc tttccggatcgatccatc	2190

<210> 86
 <211> 2190
 <212> DNA
 <213> Artificial Sequence
 5
 <220>
 <223> An A. tumefaciens mutant.

 <400> 86

10	atggtaacga	tcattcagga	tgacggagcg	gagacctacg	agacgaaagg	cggcatccag	60
	gtcagccaa	agcgcggcc	caccgattat	gccaacgcca	tcgataatta	catgaaaag	
	cttggattccc	atcgccggcg	ggtttttcg	tcaactatg	aatatccgg	ccgttacacc	120
	cgtctggata	cgcccatcg	cgatcccg	ctcggcattt	cctgttttgg	ccgcaagatg	180
	tggatcga	cctataatgg	ccgcggc	gtgctgtcg	atttcattac	ggaaaagctg	240
15	aaggcgacac	ccgatctca	cctcgccgt	tcctcgaccc	gcccgtcg	tcttaccgtc	300
	aacgaaccgg	accgtgtt	caccaagaa	gaacgtcg	aaatcccgac	ggtcttcacc	360
	gctctcagag	ccatcg	cctttctat	tcgagccgg	attccggcat	ccgcctgttc	420
	ggtcgccttcg	gttacgtatc	ccgc	ttcgcacgcg	tcaagcttc	gtctggcggt	480
	ccggaagacc	agcgtgacat	ggtgcgtt	ctgcgc	aaatccctcg	cgttgatcac	540
20	tatccgc	aggccttgat	cgacc	gatttcgaga	aggacggcat	gacgacggac	600
	ggcaatctt	ccgacattac	cccc	tciaagacca	ccgcattacat	ccgc	660
	ggcgcattacc	gtcccgccg	atattcc	cttgcgttga	aggcaagga	aagtctccgc	720
	cgccgcgacc	tgttcggat	cg	tggacgttgc	cgaaaagcaat		780
	ccgtcg	tttcccgcc	cctgaaggc	atcaacc	ccgc	cttcttc	840
25	aatttcggcg	atcaggata	tctgtcg	gcttcgc	aaatgttcgt	ggcg	900
	ggccgtcg	tcgagac	cccgat	atca	ccgc	atccgt	960
	gccgcac	agcagat	ttt	ggcaccat	agcgcggcg	cgatccgt	1020
	atgtgtcg	acgtgc	aaactcg	tcactcg	aaatcgaaa	cgactgacc	1080
	gtcattggc	gccc	ccatcgat	tcacgc	aggacgaaatc	atcgat	1140
30	gaaggccg	tgccgcg	at	tc	tc	ccacaccgt	1200
	gtcacc	cggtgc	tttgc	cc	cc	cgatcacatc	1260
	agccgc	c	gg	cc	cc		1320
	accggc	c	gttgc	at	cc		1380
	g	gttgc	at	cc	cc		1440
	g	gttgc	at	cc	cc		1500
35	tccgc	cat	tc	tc	cc		1560
	ccat	tc	tc	cc	cc		1620
	gtgcac	tc	tc	tc	cc		1680
	ccgg	tc	tc	tc	cc		1740
	ccgg	tc	tc	tc	cc		1800
40	ctgc	ccat	tc	tc	cc		1860

91

ctgcgcggc	ttgttgtgccc	catgcacggc	aaggccttcgc	gcatccgcgt	gcttggaaacc	1920
ggccctgtct	tctccggctct	cgccaaggaa	gtcacggctcg	gtcggttacca	ttcgatcttc	1980
ggcgatcccc	ccaccctgc	cgctgtatttc	atcatcacccg	cagaagcga	ggacggcacg	2040
atcatgggc	tcgaaacacgc	caaggaacccg	gtggccgcgg	ttcagttcca	cccgaaatcg	2100
5 atcatgacgc	tcggacaggaa	cgcgggcatg	cggtatcg	agaatgtcg	ggtgcatctg	2160
acccgaaagg	cgaaagccaa	ggccggctgtga				2190
<210>	87					
<211>	2190					
10 <212>	DNA					
<213>	Artificial Sequence					
<220>						
<223>	An A. tumefaciens mutant.					
15						
<400>	87					
atggtaacgc	tcattcaggaa	tgacggagcg	gagacctacg	agacgaaagg	cggcacatccag	60
gtcagccoga	agcgccggcc	caccgattat	gccaacgcga	tcgataatta	catcgaaaag	120
cttggatcccc	atcgccggcgc	ggttttttcg	tgcaactatg	aatatccggg	ccgttacacc	180
20 cgctggata	cggccatctgt	cgatccggcc	cttcggcattt	cctgtttttgg	cgccaaagatg	240
tggatcgaa	ccctataatgg	ccggccggaa	ttgtgtgtcg	atttcattac	ggaaaaaggt	300
aaggccacac	ccgatctcac	cctccggcgt	tcctggacccc	ccggccgtcg	tcttacccgtc	360
aacgaacccg	accgggtctt	caccggaaagaa	gaaacgtcgaa	aaatccggac	ggtttttcacc	420
gcttcctagag	ccatctgtcgaa	cttcttttat	tcgacggccgg	atttcggccat	cggccgtgttc	480
25 ggtgccttcg	tttacgtatc	cgccctccag	tttcgacggc	ttaaggcttcc	gcttggcgctg	540
ccggaaagacc	agcggtacat	gggtgtgttt	ctggccgatg	aaatctctgt	cgttgtatcac	600
tatccgcga	aggcctggat	cgaccgtac	gatttcaga	aggacggcat	gacgacggac	660
ggcaaattct	ccgacattac	cccegatccc	ttcaagacca	ccgataccat	cccccggaaag	720
ggcgatcacc	gtccggccga	atattccgag	cttgggtgt	aggccaaaggaa	aagtttccgc	780
30 cggggggacc	tttttcgggc	cagaattca	ttggagcggt	cgaaaaagcaat		840
ccgtcgccga	tttcccgccg	cctgaaggcg	atcaacccgt	cgcccttattc	ttttttcatc	900
aatctccggc	atcaggaata	tctggtcggc	gcctcgccgg	aaatgttgcgt	gcccgtctcc	960
ggccgtcgca	tcgagacactg	cccgatatac	ggcaccatca	agcgccggcg	cgatccgatt	1020
gcccacagcg	agcagatttt	gaaactgtctc	aactctgaaaa	aggacgaaatc	cgaaactgacc	1080
35 atgtgttcgg	acgtgtggacc	caacgaaacg	agccgcgtct	ggcggccggg	ttccgttgaa	1140
gtcattggcc	ccggcccgat	cgagatgtat	tcacggctca	tccacacccgt	cgatcacatc	1200
gaaggccgcgt	tcggcgacga	tatggacgccc	tttgcacgggt	tcctcggcc	ccggccggcc	1260
gtcaacggta	ccgggtcgacc	aaacgtgtgg	gcccgtcgat	tcatcgaa	ttatcgaaaa	1320
agccggccgc	cctgttattgg	cggtgcgtac	ggcatggtcg	gcttcaacgg	cgacatgtaa	1380
40 acccgqcccta	cqctggccac	catccggatc	aaqqacgqgtt	ttqccqaaqt	qcqccqccqcc	1440

92

gacccctgc tcaatgattc caaccgcag gaagaagaag ccggaaaccgc actgaaggcc	1500
tccggcatg tatcagccat tcgtgacgc aaaggccaca actctgccc caccaaggct	1560
gatgccgcca aagtccggcac cgccgtcaag atctctgtcg tcgaccacga agacagcttc	1620
gtgcacacgc tggcgaattt ttccgcacag acggggcgcga cggctctgac cgtcagatca	1680
5 cccggcgtcgcg ccgacgttgtt cgatcgttcc cagccggacc tgctttgtct gtccggccggaa	1740
ccccggcaggcc cgacggattt cgactgcoaa gcaacgtaca aggccggcccg cgccggcgat	1800
ctggccgatcc tggcggtttt cctcggtctg caggccattgg cagaaggctta tggccggcag	1860
ctggccgacgc ttgtctgtcc catgcacggc aagcccttcgc gcatccggcgt gttggaaecc	1920
ggccctgtct tctccggctt cggcaaggaa gtacccgtcg gtctgttacca ttgcatcttc	1980
10 gccgatcccc ccaccctgcg cgctgttattc atcatcacccg cagaaggccg ggcacccgcacg	2040
atcatggccat tgcgaaacacgc caaggaaacct ggccggcccg ttctgttcca cccggaaatcg	2100
atcatgacgc tccggacaggca cggccggcatg cggatgtatcg agaatgtcg ggtgcatctg	2160
accggcaaggcc gggccggatgtt	2190

```
15 <210> 88  
      <211> 2190  
      <212> DNA  
      <213> Artificial Sequence
```

20 <220>
<223> An *N. tumefaciens* mutant

```

<400> 88
atggtaacga tcattcagga tgacggagcg gagacctaaag agacgaaagg cggcatccag 60
25 gtcagccgaa agegcgggcc cacogattt gccaacgcga tcgataatta catcgaaaag 120
cttgattccc atcgccggcgc gggttttgc tccttctatg aataccggg ccgttacacc 180
cgctggata cgccatcgatcgatcgccgc tcggcattt ctgttttgg ccgcaagatg 240
tggatcgaa cctataatgg ccggggcgaa gtgtgtcg attcattac ggaaaaagctg 300
aaggcacac ccgatctcac cctcggtcgat tcctcgaccc gccggctcg tcttaccgtc 360
30 aacgaaccgg accgtgttcc caccgaagaa gaaacgtcgaa aatcccgac ggttccacc 420
gtctcagag ccatcgatcgatcgatccat tcgacgcgcg attcgccat ccggctgttc 480
ggtcgcctcg gttacgatct ccgttccatg ttgcacgcga tcaagcttgc gtggcggt 540
ccggaaagacc agcgtgatcat ggtgtgtt ccgtcccgatg aaatccgtcg ctgtgtac 600
tatccgcaca aggccgtggat cgacgttac gattcgaga aggacggcat gacgacggac 660
35 ggcaaatctt ccgacattac ccccgatccc ttcagacca ccgatccat ccggccaaag 720
gggatcacc gtcccgcgat atatcccgat ctgtgtgtga aggccaaaggaa attatccgc 780
cgccggaccc tggatcgatcgatccatgc cagaatccatc tgagccgtt cgaaagacat 840
ccgtccggca ttcccgcccg ctgttcccgat aatcccgat ccggccatccatccatc 900
aattcccgatccatcgatccatc ttcgtgttgc gcttcccgat aatgttccgt ccggcttcc 960
40 ggcgttccatc tcggatccatc cccatccatc ccgttccatc ccgttccatc ccgttccatc 1020

```

93

gcccacagcg	agcagattt	gaaactgctc	aactcgaaaa	aggacgaatc	cgaactgacc	1080	
atgtgctcg	acgtggaccg	caacgacaag	agccgcgtct	gcgagccggg	ttcggtgaag	1140	
gtcattggcc	gcccgcagat	cgagatgtat	tcacgcctca	tcacacccgt	cgatcacatc	1200	
gaaggccg	tcgcgcgacg	tatggacgcc	tttgacgggt	tcctcagcca	cgccctgggg	1260	
5	gtcacccgtca	ccgggtgcacc	aaagctgtgg	gccatcgct	tcatcgaaagg	tcatgaaaag	1320
agccgcgcg	cctggtatgg	cggtgcgatc	ggcatgtcg	gcttcaacgg	cgacatgaat	1380	
acccgcctca	cgctgcgcac	catccggatc	aaggacggtta	ttgccaaga	gcgcgcggc	1440	
gacccctcg	tcaatgatcc	caaccccgag	gaagaagaag	ccgaaacccg	actgaaggcc	1500	
tcgcctatc	tatcageccat	tcgtgacgca	aaaggacccca	actctgcgc	caccaagcgt	1560	
10	gatggccca	aagtggccac	cggcgtcaag	atccctgtcg	tegaccaega	agacatgttc	1620
gtgcacacgc	ttggcaattt	ttccggccag	acggggcgca	cggtctcgac	cgtcagatca	1680	
ccggcgtcg	ccgacgtgtt	cgatcgcttc	cagccggacc	tcgttgttct	gtcgccegg	1740	
ccggcagcc	cgacggattt	cgactgtcaag	gcaacgatca	aggccgcgg	cgcccccgat	1800	
ctgcgcgatc	tcggcggtt	cctcggtctg	caggatcttgc	cagaaggcta	ttggcgccgag	1860	
15	ctgogccagc	ttgtgtgtcc	catgcacggc	aaggcttcgc	gcacccgcgt	gttggaaacc	1920
ggcctegtc	tctccggctt	cggtcaaggaa	gtcacggtc	gtcggttacca	ttcgatcttc	1980	
gcccgcgttc	ccacccctgc	cggtgatttc	atcatcaccc	cagaaggcg	ggacggcacc	2040	
atcatgggc	tgcgacacgc	caaggaaaccg	gtggccgcgg	ttcagttcca	cccgaaatcg	2100	
atcatgacgc	tcggacacgga	cggggccatgc	cggtatgc	agaatgtcg	gttgcatcg	2160	
20	acccgcaagg	cgaaagaccaa	ggccgcgtga			2190	

<210> 89

<211> 2190

<212> DNA

25 <213> Artificial Sequence

<220>

<223> An A. tumefaciens mutant.

30 <400> 89

atgttaacga	tcattcaggga	tgacggagcg	gagacctacg	agacgaaagg	cgccatccag	60	
gtcagccgaa	agccgcggcc	caccgattat	gccaacgc	tcgataatta	catcgaaaag	120	
cttggattccc	atcgccggc	ggtttttctg	tccaaactatg	aatatccggg	ccgttacacc	180	
cgctgggata	cgcccatcg	cgatcgccg	cteggat	cctgttttgg	ccgcaagatg	240	
35	tggatcgaa	cctataatgg	ccgcggc	gtgtgtcg	atttcattac	ggaaaagctg	300
aaaggcgac	ccgatctc	cctccggc	tcctcg	accgc	gcccgtcg	360	
aaacgacccg	accgtgtt	cacca	gaacgtcg	aaatcccgac	ggtcttacc	420	
gctctcagag	ccatcg	cctctt	tcgac	ttccggccat	cgccatgttc	480	
40	ggtccttc	gttacgatc	cgccatccag	ttcgacg	gca	ttcgttgcgt	540
ccggaaaccc	agcgtgacat	ggtgcgtt	ctggccatcg	aaatcccgat	cggtgtatc	600	

94

tattccggca	aggcctggat	cgaccgttac	gatttcgaga	aggacggcat	gacgacggac	660
ggcaaattct	ccgcacattac	ccccatccc	ttcaagacca	ccgataccat	cccgccaaag	720
ggcgtacacc	gtcccgcgta	atattccgag	cttgggttga	aggccaagga	aagcttccgc	780
cgcggcggacc	tgttccgggt	cgttccggc	cagaaaatca	tggagcggtg	cgaaaacaaat	840
5 cgcgtccggca	tttcccgccg	cctgaaggcg	atcaacgcgt	cgcctatcc	cttcttcatc	900
aatctcgccg	atcaggaaat	tctggtcggc	gcctcgccg	aaatgttgcgt	gcccgtctcc	960
ggcgtcgca	tgcgacactg	cccgatataca	ggcaccatca	agcgcggcg	cgatccgatt	1020
gcgcacggcg	agcagatctt	gaaactgcgtc	aactcgaaaa	aggacgatac	cgaactgcacc	1080
atgtgtccgg	acgtggaccy	caacgacaag	agcgcgtct	gcccgggg	ttcggtgaaag	1140
10 gtatggcc	gcccggat	cgagatgttat	tcacgcctca	tccacacccgt	cgatcacatc	1200
gaaggccgc	tgcgcgacga	tatggacgcc	tttgacgggt	tcctcagcc	ccgcctggggcc	1260
gtcaccgtca	ccgggtgcacc	aaaactgtgg	gcccgtgcgt	tcategaagg	tcatgaaaaag	1320
agcccgccgc	cctggatattg	cggtgcgtac	ggcatgtgc	gcttcaacgg	cgacatgaat	1380
acccgcctca	cgctgcgcac	atccggatc	aaggacggtt	ttgcgcgaaat	gcccggccgc	1440
15 ggcaccttc	tcaatgtatc	caaccccgac	gaagaagaag	ccgaaacccg	actgaaggcc	1500
tccgccccat	gtatccatgc	tctggacgc	aaaggacca	actctggccgc	caccaacgg	1560
gatggccca	aaatcgccac	cggcgtcaag	atccctgtct	tgcaccacga	agacagatcc	1620
gtgcacacgc	tggcgaatta	tttccggcc	acggccgcga	cggtctcgac	cgtcagatca	1680
ccggtegcag	ccgcacgttt	cgatcgcttc	cagccggacc	tgcgttgtct	gtgcggccgg	1740
20 cccggcagcc	cgacggattt	cgactgcac	gcaacgcata	aggccggccgc	egcccgccgat	1800
ctgcgcgat	tccggcggtt	cctcggttgc	caggcatgg	cagaaggccta	tggccggcgag	1860
ctgcgcgcage	ttgtctgtgc	catgcacgc	aaggccctcg	gcatcccggt	gtctggaaaccc	1920
ggctctgtet	tctccggct	cggaaggaa	gtacacgggt	gtcggttacca	ttcgatcttc	1980
gcgcgatcccg	ccacccgtcc	gctgtatttc	atcatccac	cagaagcga	ggacggccac	2040
25 catatggca	tgcacacgc	caaggaacc	gtggccgcgc	ttcagtttca	cccgaaatcg	2100
atcatgacgc	tgcgacagga	cgccggcat	cggtatgc	agaatgtcgt	gggtgcatctc	2160
acccgcgaagg	cgaagaccaa	ggccgcgtga				2190

<210> 90

30 <211> 2190

<212> DNA

<213> Artificial Sequence

<220>

35 <223> An A. tumefaciens mutant.

<400> 90

atggtaacga	tcatttcggaa	tgacggagcg	gagacccatcg	agacgaaagg	cgccatccag	60
gtcagccgaa	agcgcggcc	caccgattat	gccaacgc	tgcataatta	catcgaaaaag	120
40 cttgatcccc	atcgccggc	ggttttttcg	tccaaactatg	aatatccggg	ccgttacacc	180

95

cgcgtggata	cgcccatcg	cgatccgccc	ctcggcattt	cctgttttg	ccgcaagatg	240	
tggatcgaa	gatcgatcg	cctataatgg	ccggggcga	gtgctgtcg	atttcattac	ggaaaagctg	300
aaggcgacac	ccatcgatc	cctcgccgt	tcctcgacc	ccgggtcg	tcttaccgtc	360	
aacgaaccgg	accgtgttt	caccaaga	gaacgtcg	aatccccac	ggttttcac	420	
5	gtcttcagag	ccatcgatcg	ccttttat	tcgagccgg	attccggat	ccggctgttc	480
ggtgccttcg	gtta	cgccatcg	ttcgacgca	tcaagcttc	gttggcggt	540	
ccggaa	accgtgtat	gttgcgttt	ctggccatg	aaatccctgt	cggtgtatc	600	
tatccgc	aggccgtgt	cgacccgtt	gatttcgaga	aggacggat	gacgacggac	660	
gccaat	cccgatccat	ccccatccc	ttcaagacca	ccgataccat	ccgccecaag	720	
10	ggcgttcc	gttgcgtat	tttgcgttgc	aggccaagg	aagcttccgc	780	
cgccggacc	gttgcgttgc	cgatccgc	ttggacgttgc	cgaaagcaat	840		
ccgtcgccg	tttccgcgc	cgttgcgttgc	atcaacgggt	cgccatattc	cttcttcata	900	
aatctcg	atcaggata	tctggtcgc	gcctcgccg	aatgttgc	gcttcgttcc	960	
ggcgtcgca	tcgagacccgt	ccggatata	ggcaccatca	agcggccgc	cgatccgtt	1020	
15	ggcgcageg	agcagat	tttgcgttgc	aactcgaaa	aggacgaaatc	cgaaactgacc	1080
atgtgtcg	acgtggacc	caacgaca	agccgcgtt	gctggccgg	ttcgggttgc	1140	
gtcattggcc	ccgcgc	cgatgttat	tcacgcctca	tccacacccgt	cgatcacatc	1200	
gaaggccgc	tgcgcgac	tatggacgc	tttgcgttgc	tcctcagcca	ccgcctggcc	1260	
gtcacccgtca	ccgggtgcacc	aaagctgttg	gcatgcgttgc	tcatcgaa	tcatgaaaag	1320	
20	aggccgcgcg	cctgttatgg	cggtgcgttgc	ggcatgttgc	gttcaacccg	cgacatgaat	1380
accggccttgc	cgctcgccac	catccggatc	aaggacggttgc	ttgcgaaatg	ggccgcgcgc	1440	
ggcacccctgc	tcaatgtatc	caacccgc	gaagaagaa	ccgaaacccg	actgaaggcc	1500	
tccgcctat	tatcgcgtat	tcgtgcgc	aaaggacca	actctcgcc	caccaagcg	1560	
gtgcgcgc	aaatcgccat	tcggcgttgc	atccgttgc	tcgaccacga	agacagcttc	1620	
25	gtcacacgc	tggcgaat	tttccgcgc	acggggccgc	cgggttcgc	cgatcgatca	1680
ccggcgtcg	ccgcgtgtt	cgatcgctt	cagccggacc	tcgttgttct	gtgcgcgc	1740	
ccggcgcgc	cgacggattt	cgactcgac	gcaacgatca	aggccgcgc	cgcccgcgat	1800	
ctgcgcgtt	tccgcgttgc	cgttgcgttgc	cggatattgc	cagaacgc	tggcgcgc	1860	
ctgcgcgc	ttgtgtgc	catgcacgc	aaatccgtc	gcatccgcgt	gttggaaatcc	1920	
30	ggctcgat	tctccgttgc	cggaaggaa	gtcacgttgc	gtcgat	tccat	1980
ccgcgcgc	ccacccgttgc	cggtgtattt	atcatcacc	cagaagcg	ggacggccac	2040	
atcatggca	tgcgacacgc	caaggaccc	gtggccgc	ttcgttcca	cccgaaatcg	2100	
atcatgacgc	tcggacagg	cgccggat	cggtatgttgc	agaatgttgc	gttgcgtatcg	2160	
acccgc	cgaa	ccaa	ggccgcgttgc			2190	

35

<210> 91

<211> 2190

<212> DNA

<213> Artificial Sequence

40

<220>

<223> An A. tumefaciens mutant.

<400> 91

5 atggtaacga tcattcagga tgacggagcg gagacctacg agacgaaagg cggcatccag	60
gtcagccgaa agcgcggccc caccgattt gcacaacgcca tcgataatta catcgaaaag	120
cttgattccc atcgcggcgc ggtttttcg tccaaactatg aatatccggg cggttacacc	180
cgtcgggata cgccgatctgt cgatccggcg ctggcattt cctgttttg cccgaagatg	240
tggatcgaag cctataatccg ccggcgccga gtgcgtctgc atttcattac gaaaaagctg	300
10 aaggcgacac cccatctccat ccccgccgtc ttctcgaccc gccggctcga tcttaccgtc	360
aacgaaccgg accgtgttcc caccgaagaa gaacgtcga aatcccgac ggtcttcacc	420
gtcttcagag ccatagtcga cttttcttat tcgaggccgg attcggccat cggctgttc	480
ggtccttcg gttaacgtat cgccttcag ttgcgcgcga tcaagcttc gctggcggt	540
ccggaaagcc acgtgtacat gttgtgttt ctgcggatg aaatccctgt cgttgatcac	600
15 tattccgcca aggccgttat cgcacgttac gatttcgaga aggacggcat gacgacggac	660
ggcaaattcc cccgacattac ccccgatccc ttcaagacca ccgataccat cccggccaaag	720
ggcgtacacc gtcccgccga atattccgag ctgtgttgta aggccaagga aagcttccgc	780
cggcgccaccc tttcccgccg cagaaattca ttggacgttg cggaaagcaat	840
ccgtcgccga ttttcccgccg ctgtaaaggcg atcaacccgt cgccttattc ctgggttcatc	900
20 aatctcgccgatc caccgaata tctggtcggc gcttcgcggc aaatgttctgt ggcgtctcc	960
ggccgtcga tcgagacatcg cccgatatca ggacccatca agcgcggcga cgatccgatt	1020
ggccacagecg agcgatattt gaaactgtc aactcgaaaaa aggacgaaatc cgaactgacc	1080
atgtgttcgg acgtggaccc caacgacaaag agcgcgtct gcgcggccggg ttccgttgaag	1140
gtcatttggcc ggcgcagat cgagatgtat tcacgcttc tccacacccgt cgatcacatc	1200
25 gaaggccgccc tgccgcacgtat tggtacgcg tttgcgggtt tccttcgcga cgcctggggc	1260
gtcacccgtca cccgtgcacc aaagctgtgg gccatcggtt tcatacgaaagg tcataaaaaag	1320
agcccgccgat cctggatattt cgggtgcgatc ggcatggcgc gcttcacacgg cgacatgaat	1380
accggccgtca cgtcgccac catccggatc aagacgcgtt ttgcgcgaat ggcgcggccgc	1440
ggcacccgtc tcaatgtatc caaccccgacg gaagaagaag ccgaaacccgactgaaggcc	1500
30 tccgcctatga tatacgccat tctgtacgcgaa aaaggacccaa actctggccgc caccacgtt	1560
gatggccca aagtccggc acggcgatccaaatccgtctcg tcgaccacgcg agacatgttc	1620
gtgcacacgcg tggcaatttttccgcac acggccgcgcg cggctctgcac cgtcgcacatca	1680
ccggctcgccatcg cgcacgttgcgtt cgcacgttgc cagccggacc tctgtgttctt gtcgcggca	1740
cccgccagcc cgacggattt cgactgtcaag gcaacgtatca aggccgcggcccg cgcggccatcg	1800
35 ctggccatcgat ttcggcggtt ctccgtgtcg caggccatgg cagaacgtatca tggccggcgag	1860
ctggccatcgatccatccgcgc aaggccgttcgc gcacccgcgt gtcgttaccc	1920
ggccgttcgtt ttcggcgatccatccgcgc aaggccgttcgc gcacccgcgt gtcgttaccc	1980
ggccgttcgtt ttcggcgatccatccgcgc aaggccgttcgc gcacccgcgt gtcgttaccc	2040
atcatggccatccatccgcgc aaggccgttcgc gcacccgcgt gtcgttaccc	2100
40 atcatgtatccatccgcgc aaggccgttcgc gcacccgcgt gtcgttaccc	2160

97

```

acccgcgaagg cgaagaccaa ggccgcgtga 2190

<210> 92
<211> 2190
5 <212> DNA
<213> Artificial Sequence

<220>
<223> An A. tumefaciens mutant.

0
<400> 92
atggtaacga tcattcagga tgacggagcg gagacacctacg agacgaaagg cggcatccag 60
gtcagccgaa agcgcgggcc caccgattt gccaacgcca tgcataatta catgaaaag 120
cttgattccc atcgccggcgc gggttttaag tccaactatg aataccggg ecgttacacc 180
5 cgctgggata cggccatcg tgcattccgc ctccggcattt cctgttttg ccgaagatg 240
tggatcgaag cctataatgg cccggcgaa gtgtgtctcg atttcattac ggaaaagctg 300
aaggcgacac cgcatttcac cctccggcgt tccctcgaccc gcccgtcga tcttaccgtc 360
aacgaacccg accgtgtctt caccgaagaa gaacgctcga aaatcccgac ggttccatc 420
0 gctctcgatg ccattcgatc ctttccatc tccagccgttccatc cggccgttc 480
gggtccatcg gttacgatc ccccttcac ttccggcgca tcaagcttc gtcgtccgt 540
ccgaaagacc aegcgatcat ggtgttccatc tccggccatg aaatccctcgatc 600
tatcccgca aggactcgat ggcggatcattt gatccggaga aggacggcat gacgacggac 660
ggcaaatctt ccgcattac ccccgatccc ttcaagacca ccgcattatc cccggccaaag 720
ggcgatcacc ttccggcga atttcggat cttgtgttccatc tccggccatc 780
5 cccggccacc ttccggat cttccggcga cagaaatcga tggagcggtt ccggaaatc 840
ccgtcgccatc ttccggccatc cttccggcga atcaacccgt ccccttccatc 900
aatctcgccatc tccggccatc ttccggccatc cttccggcga atcaacccgt ccccttccatc 960
ggccgtccatc tccggccatc cttccggccatc cttccggccatc cttccggccatc 1020
ccgcacatcg agcagatttt gaaactcgatc aactcgaaaa aggacgaaatc cgaactcgacc 1080
0 atgtgtctgg acgtggaccat cttccggccatc cttccggccatc cttccggccatc 1140
gtcattggcc tccggccatc cttccggccatc cttccggccatc cttccggccatc 1200
ccgcacatcg agcagatttt gaaactcgatc aactcgaaaa aggacgaaatc cgaactcgacc 1260
gtcaccgtca cttccggccatc cttccggccatc cttccggccatc cttccggccatc 1320
ccgcacatcg agcagatttt gaaactcgatc aactcgaaaa aggacgaaatc cgaactcgacc 1380
5 acggccgtca cttccggccatc cttccggccatc cttccggccatc cttccggccatc 1440
ccgcacatcg agcagatttt gaaactcgatc aactcgaaaa aggacgaaatc cgaactcgacc 1500
gtcaccgtca cttccggccatc cttccggccatc cttccggccatc cttccggccatc 1560
ccgcacatcg agcagatttt gaaactcgatc aactcgaaaa aggacgaaatc cgaactcgacc 1620
gtcaccgtca cttccggccatc cttccggccatc cttccggccatc cttccggccatc 1680
0 ccgcacatcg agcagatttt gaaactcgatc aactcgaaaa aggacgaaatc cgaactcgacc 1740

```

ccggcagcc	cgacggatt	cgactgcaag	gcaacgatca	aggccgcgg	egccccgcat	1800
ctgcccata	tccgggttg	cctcggtctg	caggcatttg	cagaaggctca	tggcgccgag	1860
ctgcccacgc	tttgtgtgcc	catcacgcgc	aaggcttcgc	gcateccgct	gctggaaacc	1920
ggctcgatct	tctccggctc	cggcaaggaa	gtcacggtcg	gtcggttacca	ttcgatctc	1980
5 gccgatcccg	ccaccacctgc	gcgtgatttc	atcatcaccc	cagaaaggcga	ggacggcactg	2040
atcatgggc	tgcgaaacacgc	caagggaaacc	gtggccgcgg	ttcagttcca	cccgaaatcg	2100
atcatgacgc	tccggacaggaa	cggatcgatcg	cgaaatgtcg	ggtgcgtatcg		2160
acccgcaagg	cgaaagaccaa	ggccgcgtga				2190

10 <210> 93

<211> 2190

<212> DNA

<213> Artificial Sequence

15 <220>

<223> An A. tumefaciens mutant.

<400> 93						
atggtaacga	tcattcagga	tgacggagcg	gagacctacg	agacgaaagg	cgccatccag	60
20 gtcagccaa	agcgcggcc	caccgattat	gccaacgcca	tcgataatta	catcgaaaag	120
cttggatcccc	atcgccgcgc	ggttttttcg	tccaaatctatg	aatatccggg	cggttacacc	180
cgtgggata	cgccatcg	cgatccgcgc	ctccggattt	cctgttttgg	ccgaagatg	240
tggatcgaa	cctataatgg	ccggggcggaa	gtgtcgatcg	atttcattac	ggaaaagctg	300
aaggccgacac	cgcgtatccac	cctcgatcgat	tcgtcgatcg	gcccgttgcg	tcttacccgtc	360
25 aacggaaacccg	acctgttctt	cacggaaaggaa	gaacgtcgaa	aaatcccgac	ggttccatcc	420
gtcttcagag	ccatcgatcg	ccttttttat	tcgagccgcgc	attccggccat	cgccgtgttc	480
ggtagcttcg	gttacatct	cgccctccag	ttcgagccgcg	tcaagtttc	gctggcggt	540
ccggaaagacc	agcggtgatct	gggtgtgttt	ctggccgtatg	aaatccctcg	cggtgtatcac	600
tattccgcga	aggcgtgtat	cgaccgttac	gatttcgaga	aggacggcat	gacgacggac	660
30 ggccaaatctt	ccgacattac	ccccgatccc	ttcaagacca	ccgatccat	ccgcggcaag	720
ggcgatcacc	gtcccgccg	atattccgag	cttgggtgt	aggccaaagg	aagttccgc	780
cgccggacc	tgttcggatgt	cggtcccgcc	cagaaaattca	tggagcggtt	cgaaaagcaat	840
ccgtccggca	tttcccgccg	cctgaaggcg	atcaaccctgt	cgccctatcc	cgccctatcc	900
aatctccggcg	atcaggata	tctggtcggc	gcttcggccg	aatgttccgt	gcccgtctcc	960
35 ggcgcgtcg	tcgagacccgt	cccgatata	ggcaccatca	agcgccggcg	cgatccgtt	1020
ggccacacgg	agcagatttt	gaaactgtct	aactcgaaaa	aggacgcatac	cgaaactgacc	1080
atgtgtcg	acgtggacccg	caacgacacg	agccgcgtct	gcaagccggg	ttcgggtgaag	1140
gtcattggcc	ggccgcagat	cgagatgtat	tcacgcctca	tccacacccgt	cgatcacatc	1200
gaaggccgc	tgcgcgcgac	tatggacgccc	tttgacgggtt	tcctcagcca	ccgcctggcc	1260
40 gtcaccgtca	ccgggtgcacc	aaagctgtgg	gccatcgatcg	tcatacgaaagg	tcatgaaaag	1320

99

agccccgcgc	cctggtatgg	cggtgcgatc	ggcatggtc	gcttcaacgg	cgacatgaat	1380
acggcctga	cgctgcgcac	catccggatc	aaggacggta	ttgccgaagt	gcgcgcggc	1440
gacccctgc	tcaatgattc	caaccgcag	gaagaagaag	cgaaaacgta	actgaaggcc	1500
tcccccattt	tatcagccat	tcgtgacca	aaaggcacca	actctgcgc	caccaaggt	1560
5 gatggccca	aagtggcac	cggcgtcaag	atccctgtcg	tcgaccacga	agacagcttc	1620
gtgcacacgc	ttggcgaatta	ttccggccag	acggggcga	cgtctcgac	cgtcagatca	1680
ccggcgcag	cgacgtgtt	cgatcgctc	cagccgacc	tcgttgtct	gtgcgcggga	1740
cccgacggc	cgacgattt	cgactgcaag	gcaacgatca	agggccccc	cgcccgcgat	1800
ctgcgcgtat	tcggcggtt	cctcggtctg	caggcatttg	cagaagctta	ttggcgcgag	1860
10 ctgcgcgcage	tttgtgtgcc	catgcacggc	aaggcttcgc	gcatccgcgt	gttggAACCC	1920
ggcctgtgt	tctccgtgt	cgccaaggaa	gtcacggcg	gtcggttacca	ttcgatcttc	1980
gccccatccgg	ccacccgtcc	gggtgtattt	atcatcaccc	cagaagcgta	ggacggcaca	2040
atcatgggc	tcgaacacgc	caaggaaacc	gtggccgcgg	tttagttcca	cccgaaatcg	2100
atcatgacgc	tcggacaggc	cgccggcatg	cggatgatcg	agaatgtcg	ggtgcatactg	2160
15 acccgcaagg	cgaagaccaa	ggccgcgtgt				2190

<210> 94

<211> 1821

<212> DNA

20 <213> Oryza sativa

<400> 94

atggagttcca	tcggccgcgc	cacgttcaacg	ccctcgccgc	tcggccgcgg	ccccgcacact	60
ccggccgggg	cgggggcccc	ggttagagcg	agggggggg	tagggcagg	aggggagggg	120
25 aggaegagta	ggcgcggccgg	cgtgagggtgc	tccgggggg	agccagaggc	aaggcgcgtg	180
atcaacgggg	gcccggccgc	gggggggggg	gaggaggaca	ggaggcgttt	tttcgaggcg	240
gccccgggtt	ggagcggggaa	gggcaacctg	gtgcccattt	gggagtgcat	cgtctccgac	300
cacctcaccc	ccgtgtcgcc	ctaccgtgc	ctcgcccc	aggacacacat	ggagacgccc	360
agtttctctt	tcgagtcgt	cgagcgggg	cccgaggggc	ccaccaacgt	cggtcgat	420
30 agcatgggtt	gaggccaccc	agtgtatgg	gtcgtggca	aggagcacaa	ggtcacaatc	480
atggaccacg	agaaggggca	gtgtacgg	cagggtcg	atgtatctat	gcatatcccc	540
aggagcatga	ttggaaaggat	gcaccccgac	cagatcgatc	agtcctccga	ttcccttcacc	600
gggtgtatgg	tcgggttctt	ttccatgtat	acagtcgtt	atgttggaaa	gaagaagctg	660
cccttcctccg	gtgtccccc	ggacgtatagg	aacccctctg	atgttccact	ttgggtttat	720
35 gatgtatgtt	tcgttcttcg	caatgtcgag	aagaaatgtat	atgtatccca	ttgggttaat	780
cttgcgtccgc	atgaacccac	cgaggatgtca	ttccaatgtat	gcaagtcccg	gctgacactg	840
ttgtatctta	aagtgcacaa	ttcaatgtat	cccaactttt	ctccaggattt	tgtaaagtt	900
cacactcgcc	agtttggta	acccttgaac	aaatcaacca	tgacaatgt	tgagtacaag	960
40 agccagagggt	tttgagaggc	gacatacgcc	aatccatgtt	aagtctatcg	agctttacga	1020
						1080

100

15 <210> 95

52113-1498

<313> DNA

5213a *Oryza sativa*

20-€400> 95

101

cgccgcggcca	cgttccagca	gctcgacgcc	gtcgggtgagt	ctccgtatca	aatgtggggg	1260
ggcatgttctt	ggtttgcggg	ttgggggtt	gatttgaatg	tgtgttctcg	tgccgcagc	1320
ggtagggag	gaggagtcca	agttcaaggc	ggggccggcg	gagggttgc	acatcctgc	1380
gctcaagcga	tgcatcttc	ccgaccacct	cacgcgggtg	ctcgcttacc	gctgccttgt	1440
5	caggaggac	gaccgcgagg	cggccagett	cctgtttgag	tccgtcgagc	1498

<210> 96

<211> 2073

<212> DNA

10 <213> Zea mays

<400> 96

gaattccggcc	aaatcgggct	atacatatcaa	cgtcgact	tagggagcgt	gaagccagcg	60	
gcaatggaa	ttcccttagccg	ccacccctcg	gttcggcc	ttcccgctcg	ccgtccccgc	120	
15	ggccggccgc	ctgggttggg	ccccggacgg	ggtaccaacc	aggccgacga	gagccggag	180
cggaaaccaggc	gggtgttaat	gctctgtgc	cgtgacgcgc	caggcgagcc	cagtgattag	240	
caggagcgct	ggccggccgc	aggccggcgga	ggaggacaa	aggccgttct	tcgaggccgc	300	
ggccgggggg	aggccggaaagg	ggaaacctgtt	gccccatgtt	gagtgcata	agggaaacct	360	
ggtccccatg	ttgggggtgc	tcgtgtcgga	ccatcttacc	cccgtgtcg	cctaccgcgt	420	
20	cctgtgtcccc	gaggacaacg	tcgacgcggcc	cagtttctc	ttcgagtcg	tcgagccagg	480
gccccaggcc	accaccaacg	tcggccgcta	tagcatgtt	ggagccacc	cagtgtatgg	540	
gattgtggcc	aaagacccaa	aggttatcat	catggaccac	gagaagagcc	aagtgcacaa	600	
gcaggatgtg	gacgaccgg	tgcagatccc	gaggacat	atggggat	ggcacccaca	660	
gcagatcgac	gagtcctctc	aatcccttc	cgggtggat	tttgggttct	tttcttatgt	720	
25	tacgggttag	tatgttggaa	agaagaagct	accgttttcc	agtgttctc	aggacgtat	780
gaacccccc	gtatgttcc	gtatgttcc	tgggactcta	tgtatgtt	ctagtcttc	ataatgttga	840
gaagaaatgt	tatgttatacc	atggggtaa	tgtggacccgg	catgcatttgc	tttggggat	900	
ataccaaatgt	ggcaggccccc	gactaaacat	tttgcataat	aaatgtcaca	tttccatgt	960	
cccacccactc	tctcttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	1020	
30	caagtcgacc	atgacaatgt	atgagtataa	aatgtgtt	cttcggcc	aggacatata	1080
tatggctggg	gatatcttc	agattttttt	aaggccagg	tttcggcc	aggacatata	1140	
caacccat	gagggttatac	gaggattac	gattgttata	ccttcggcc	acatggcc	1200	
tgtacaggca	agaggctgt	tattttgtc	tttgcataat	tttgcataat	tttgcataat	1260	
taaggggaa	atttataatc	gaccatttc	tttgcataat	tttgcataat	tttgcataat	1320	
35	gaaatgtcaa	atgcaagacg	agcaactgtt	aaatgtatgg	aaacatgttgc	ccgacacat	1380
aatgttttgc	gacttggaa	gaaatgtatgt	tttgcataat	tttgcataat	tttgcataat	1440	
gaagggtgg	aaatgttata	tttgcataat	tttgcataat	tttgcataat	tttgcataat	1500	
tagtggac	tttgcataat	tttgcataat	tttgcataat	tttgcataat	tttgcataat	1560	
tggacatgt	tttgcataat	tttgcataat	tttgcataat	tttgcataat	tttgcataat	1620	
40	tacggggcga	ggaccatata	tttgcataat	tttgcataat	tttgcataat	tttgcataat	1680

102

aattgcactt	tctctccgca	ccatcgatt	ctcaacagcg	ccgagccaca	acacgatgt	1740
ctcataaaaa	gacgcgata	ggcgctggga	tggggtcgct	catcttcagg	ctggcgcagg	1800
cattgttgcc	gacatgcgc	cagatgacga	acaacgtgaa	tgcgagaata	aggctgtgc	1860
actagtcgg	gccccatgc	ttgcgagatc	agcttttgt	aaaaaagaat	agtgtgtat	1920
5 gggttatcggt	tagttctgt	tcatgtttct	tttaccact	ttccgttaaa	aaaagatgtc	1980
attagtgggt	ggagaaaagc	aataaactg	tttcttagag	aaccgaagaa	atatggaaat	2040
tgagggtatg	gcccgaattc	ctgcgcggc	ggg			2073

<210> 97

10 <211> 504

<212> DNA

<213> *Triticum aestivum*

<400> 97

15 cccaaacagt	ggggccttag	gaggatata	atttgatggt	gacatgttta	tcgtctttgc	60
tctccgcacc	atttggtttt	caacagctcc	aagccccat	aggatgtact	catacaaaaag	120
ctcagatagg	ccccgagagt	gggttgccta	tcttcaggct	ggtgccggca	ttgttgcgtga	180
tagtatccca	gacgatgagc	aaaaaagaatt	tgagaataag	gcggctgccc	tagctcgggc	240
aattgtatctt	gcagagttgg	cttttttaga	caaagaataat	agtgtctatt	aaatttattt	300
20 tttttagtttgc	tcatcatttt	tcacccaggat	cattttggaa	agttgttcat	cgtttttca	360
ccaggttcat	attggggaaa	aaagcaata	ccgtttttgtt	gtcctttgaa	atgaataaat	420
ttgagctata	ataagatgt	ttttgtcat	cgggcaaaaa	aaaaaaaaaa	aatataaaaa	480
aaaaaaaaaa	aaaaaaaaaa	aata				504

25 <210> 98

<211> 2161

<212> DNA

<213> *Nicotiana tabacum*

30 <400> 98

gtcaaaaaat	cccatttcac	cgtttctcg	tttctcttcc	tcaactaattt	tgtctctttc	60
tcttggtttg	ctatgtgt	cttgcggaa	tgca	cgatcggt	acatatatca	120
ttccggccac	ccacccggaaa	gttctgcatt	tcgcgcgtat	ttctagccgg	agctcaactt	180
ctgcacttgc	gcttcgtgtc	cgtacactac	aatgcgcgtg	ccttcactct	tcatctctag	240
35 ttatggatga	ggacaggatc	atggaaatgtt	ctaaaaacgg	gaaacttgatt	ccgtcgacaca	300
aaaccatttt	ttctgtatcat	ctgactccgg	tgctggctta	ccgggttttg	gtgaagaag	360
acgacccgtga	agctccaage	tttctcttttgc	aatccgttgc	acctgggtttt	cgagggttctca	420
gtgttggatgc	ctacagcggt	gtgggggttc	aaccatctat	ggaaatttg	gctaaggaaac	480
acaatgtgac	tatattggac	caccacactg	gaaaattgac	ccagaagact	gtccaagatc	540
40 ccatgcacat	tccgaggaggat	atttctgagg	gatggaaac	cagactcatt	gatgaatctc	600

103

ctgataacctt ttgtgggtga tgggttggtt atttctcata tgacacagtt cggatgttag 660
 agaacaggaa gttgccattc ctaagggttc cagaggatga cggAACCTT gcagatattc 720
 aatttaggact atacgaatgt gtcattgtgt ttgtatcatgt tgagaagaaa gcacatgtga 780
 ttcactgggt gcagttggat cagtattcat ctcttcgtga ggcataatcgat gatggaaaga 840
 5 aacgcttggaa aatatttagtgc tctagact aaggaaatgtga gtcTCCAAAGG ttatccccc 900
 gttctgtggaa ttctgtact catgtttttg gacccttcatt aaccaaggaa aacatgacaa 960
 gtgaggagta caagaatgtc gtcttacaag caaaggagca cattgtgcga ggagacat 1020
 ttcaaatgtc tttaagtcaa cgctttgtgaa gaagaacatt tggtgcggca ttggaaatgt 1080
 acagagactt aagaattgtc aatccaaacgc catatatgtc ttatcacatcgac gccaggct 1140
 10 gtatTTTGT tgcatcgagc ccagaaatTTT tgacacgttgaa gagaagaga aagatTTTGA 1200
 atcgaccact ggctggagaca aggcaaggaa ggaaagacacc tgatgaggat gtgtgttgg 1260
 aaatcgatgt gttaaaatgt gagaacaacaa ggcaggagca catcatgtcg ttggattttg 1320
 gacggaaatgt tgtaggaaag gtgtcaaaac ctggTTTGT gtatgtcgaa aagctcatgaa 1380
 ggggttggatc gtatccccat gtgtgcaca taagctccac ggTCTCTGGA gagttgttg 1440
 15 atcatttaac ctgtgggtat gcactacgtg ctgcattggc tggtgggacc gtcaagtggag 1500
 cacccaaaggaa aaaggccatg gagttgattgt atcagctaga agtagctgg agaggccctt 1560
 acagttgtgg ttgtggggc atttccctttt cagggtgacat ggacatcgca ctatgtctaa 1620
 ggacgatgtt attcctcaat ggacgtcgat atgcacaat gtattcatat acagatgcac 1680
 gcaaggcgatc ggaatgggtt gtcatctcc aatccggggg tggaaatgtg gtcgtatgt 1740
 20 atccctgtatc ggaacagata gaatgcgaga ataagaatgtc cggatgtgc cgaggccattg 1800
 acttggccoga gtcaatTTTGT gtaaaggaa gacacaaacc gtcaatgtcaag ataaatgtt 1860
 ctgtggccaaatc tctatTTTCA agggttacaaatc gtcaacatc tggtatgtcg aagagcagag 1920
 tacatgtatc gaaaggaaactcg cgaatatgtaa gtgtatcataa aattctaaatg tggtttttttt 1980
 gttcgtatTTTGT ggtatgtatc tgtagtgcgt tgtagatgtt gttttagatgt 2040
 25 accttcattt tgggtttctt gaaaggacatc tgccatata tagcaatTTTGT atcaatgtt 2100
 taatccctgtt atgcgggttgc gaaatgtcaatc catcagaatc ttggaaaaaaa aaaaaaaaaaa 2160
 a

<210> 99

30 <211> 606

<212> PRT

<213> Oryza sativa

<400> 99

35 Met Glu Ser Ile Ala Ala Ala Thr Phe Thr Pro Ser Arg Leu Ala Ala

1 5 10 15

Arg Pro Ala Thr Pro Ala Ala Ala Ala Pro Val Arg Ala Arg Ala

20 25 30

Ala Val Ala Ala Gly Gly Arg Arg Arg Thr Ser Arg Arg

104

Arg Cys Ser Ala Gly Lys Pro Glu Ala Ser Ala Val Ile Asn Gly Ser
 50 55 60
 Ala Ala Ala Arg Ala Ala Glu Glu Asp Arg Arg Arg Phe Phe Glu Ala
 65 70 75 80
 5 Ala Glu Arg Gly Ser Gly Lys Gly Asn Leu Val Pro Met Trp Glu Cys
 85 90 95
 Ile Val Ser Asp His Leu Thr Pro Val Leu Ala Tyr Arg Cys Leu Val
 100 105 110
 Pro Glu Asp Asn Met Glu Thr Pro Ser Phe Leu Phe Glu Ser Val Glu
 10 115 120 125
 Gln Gly Pro Glu Gly Thr Thr Asn Val Gly Arg Tyr Ser Met Val Gly
 130 135 140
 Ala His Pro Val Met Glu Val Val Ala Lys Glu His Lys Val Thr Ile
 145 150 155 160
 15 Met Asp His Glu Lys Gly Lys Val Thr Glu Gln Val Val Asp Asp Pro
 165 170 175
 Met Gln Ile Pro Arg Ser Met Met Glu Gly Trp His Pro Gln Gln Ile
 180 185 190
 Asp Gln Leu Pro Asp Ser Phe Thr Gly Gly Trp Val Gly Phe Phe Ser
 20 195 200 205
 Tyr Asp Thr Val Arg Tyr Val Glu Lys Lys Lys Leu Pro Phe Ser Gly
 210 215 220
 Ala Pro Gln Asp Asp Arg Asn Leu Pro Asp Val His Leu Gly Leu Tyr
 225 230 235 240
 25 Asp Asp Val Leu Val Phe Asp Asn Val Glu Lys Lys Val Tyr Val Ile
 245 250 255
 His Trp Val Asn Leu Asp Arg His Ala Thr Thr Glu Asp Ala Phe Gln
 260 265 270
 Asp Gly Lys Ser Arg Leu Asn Leu Leu Leu Ser Lys Val His Asn Ser
 30 275 280 285
 Asn Val Pro Lys Leu Ser Pro Gly Phe Val Lys Leu His Thr Arg Gln
 290 295 300
 Phe Gly Thr Pro Leu Asn Lys Ser Thr Met Thr Ser Asp Glu Tyr Lys
 305 310 315 320
 35 Asn Ala Val Met Gln Ala Lys Glu His Ile Met Ala Gly Asp Ile Phe
 325 330 335
 Gln Ile Val Leu Ser Gln Arg Phe Glu Arg Gln Thr Tyr Ala Asn Pro
 340 345 350
 Phe Glu Val Tyr Arg Ala Leu Arg Ile Val Asn Pro Ser Pro Tyr Met
 40 355 360 365

105

Ala Tyr Val Gln Ala Arg Gly Cys Val Leu Val Ala Ser Ser Pro Glu
 370 375 380
 Ile Leu Thr Arg Val Arg Lys Gly Lys Ile Ile Asn Arg Pro Leu Ala
 385 390 395 400
 5 Gly Thr Val Arg Arg Gly Lys Thr Glu Lys Glu Asp Glu Met Gln Glu
 405 410 415
 Gln Gln Leu Leu Ser Asp Glu Lys Gln Cys Ala Glu His Ile Met Leu
 420 425 430
 Val Asp Leu Gly Arg Asn Asp Val Gly Lys Val Ser Lys Pro Gly Ser
 10 435 440 445
 Val Lys Val Glu Lys Leu Met Asn Ile Glu Arg Tyr Ser His Val Met
 450 455 460 .
 His Ile Ser Ser Thr Val Ser Gly Glu Leu Asp Asp His Leu Gln Ser
 465 470 475 480
 15 Trp Asp Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly Ala
 485 490 495
 Pro Lys Val Lys Ala Met Glu Leu Ile Asp Glu Leu Glu Val Thr Arg
 500 505 510
 Arg Gly Pro Tyr Ser Gly Gly Leu Gly Gly Ile Ser Phe Asp Gly Asp
 20 515 520 525
 Met Leu Ile Ala Leu Ala Leu Arg Thr Ile Val Phe Ser Thr Ala Pro
 530 535 540
 Ser His Asn Thr Met Tyr Ser Tyr Lys Asp Thr Glu Arg Arg Glu
 545 550 555 560
 25 Trp Val Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ser
 565 570 575
 Pro Asp Asp Glu Gln Arg Glu Cys Glu Asn Lys Ala Ala Leu Ala
 580 585 590
 Arg Ala Ile Asp Leu Ala Glu Ser Ala Phe Val Asp Lys Glu
 30 595 600 605

<210> 100
 <211> 67
 <212> PRT
 35 <213> Oryza sativa

<400> 100
 Met Cys Val Leu Val Ala Ala Ala Val Arg Glu Glu Ser Lys Phe
 1 5 10 15
 40 Lys Ala Gly Ala Ala Glu Gly Cys Asn Ile Leu Pro Leu Lys Arg Cys

106

20	25	30
Ile Phe Ser Asp His Leu Thr Pro Val Leu Ala Tyr Arg Cys Leu Val		
35	40	45
Arg Glu Asp Asp Arg Glu Ala Pro Ser Phe Leu Phe Glu Ser Val Glu		
5 50	55	60
Gln Gly Ser		
65		
<210> 101		
10 <211> 525		
<212> PRT		
<213> Zea mays		
<400> 101		
15 Met Trp Glu Cys Ile Lys Gly Asn Leu Val Pro Met Trp Glu Cys Ile		
1 5 10 15		
Val Ser Asp His Leu Thr Pro Val Leu Ala Tyr Arg Cys Leu Val Pro		
20 25 30		
Glu Asp Asn Val Asp Ala Pro Ser Phe Leu Phe Glu Ser Val Gln		
20 35 40 45		
Gly Pro Gln Gly Thr Thr Asn Val Gly Arg Tyr Ser Met Val Gly Ala		
50 55 60		
His Pro Val Met Glu Ile Val Ala Lys Asp His Lys Val Thr Ile Met		
65 70 75 80		
25 Asp His Glu Lys Ser Gln Val Thr Glu Gln Val Val Asp Asp Pro Met		
65 90 95		
Gln Ile Pro Arg Thr Met Met Glu Gly Trp His Pro Gln Gln Ile Asp		
100 105 110		
Glu Leu Pro Glu Ser Phe Ser Gly Gly Trp Val Gly Phe Phe Ser Tyr		
30 115 120 125		
Asp Thr Val Arg Tyr Val Glu Lys Lys Lys Leu Pro Phe Ser Ser Ala		
130 135 140		
Pro Gln Asp Asp Arg Asn Leu Pro Asp Val His Leu Gly Leu Tyr Asp		
145 150 155 160		
35 Asp Val Leu Val Phe Asp Asn Val Glu Lys Lys Val Tyr Val Ile His		
165 170 175		
Trp Val Asn Val Asp Arg His Ala Ser Val Glu Glu Ala Tyr Gln Asp		
180 185 190		
Gly Arg Ser Arg Leu Asn Met Leu Leu Ser Lys Val His Asn Ser Asn		
40 195 200 205		

107

Val Pro Thr Leu Ser Pro Gly Phe Val Lys Leu His Thr Arg Lys Phe			
210	215	220	
Gly Thr Pro Leu Asn Lys Ser Thr Met Thr Ser Asp Glu Tyr Lys Asn			
225	230	235	240
5 Ala Val Leu Gln Ala Lys Glu His Ile Met Ala Gly Asp Ile Phe Gln			
245	250	255	
Ile Val Leu Ser Gln Arg Phe Glu Arg Arg Thr Tyr Ala Asn Pro Phe			
260	265	270	
Glu Val Tyr Arg Ala Leu Arg Ile Val Asn Pro Ser Pro Tyr Met Ala			
10 275	280	285	
Tyr Val Gln Ala Arg Gly Cys Val Leu Val Ala Ser Ser Pro Glu Ile			
290	295	300	
Leu Thr Arg Val Ser Lys Gly Lys Ile Ile Asn Arg Pro Leu Ala Gly			
305	310	315	320
15 Thr Val Arg Arg Gly Lys Thr Glu Lys Glu Asp Gln Met Gln Glu Gln			
325	330	335	
Gln Leu Leu Ser Asp Glu Lys Gln Cys Ala Glu His Ile Met Leu Val			
340	345	350	
Asp Leu Gly Arg Asn Asp Val Gly Lys Val Ser Lys Pro Gly Gly Ser			
20 355	360	365	
Val Lys Val Glu Lys Leu Ile Ile Glu Arg Tyr Ser His Val Met His			
370	375	380	
Ile Ser Ser Thr Val Ser Gly Gln Leu Asp Asp His Leu Gln Ser Trp			
385	390	395	400
25 Asp Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly Ala Pro			
405	410	415	
Lys Val Lys Ala Met Glu Leu Ile Asp Lys Leu Glu Val Thr Arg Arg			
420	425	430	
Gly Pro Tyr Ser Gly Gly Leu Gly Gly Ile Ser Phe Asp Gly Asp Met			
30 435	440	445	
Gln Ile Ala Leu Ser Leu Arg Thr Ile Val Phe Ser Thr Ala Pro Ser			
450	455	460	
His Asn Thr Met Tyr Ser Tyr Lys Asp Ala Asp Arg Arg Glu Trp			
465	470	475	480
35 Val Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ser Pro			
485	490	495	
Asp Asp Glu Gln Arg Glu Cys Glu Asn Lys Ala Ala Ala Leu Ala Arg			
500	505	510	
Ala Ile Asp Leu Ala Glu Ser Ala Phe Val Asn Lys Glu			
40 515	520	525	

<210> 102
<211> 92
<212> PRT
5 <213> *Triticum aestivum*

<400> 102
Pro Asn Ser Gly Gly Leu Gly Gly Ile Ser Phe Asp Gly Asp Met Leu
1 5 10 15
10 Ile Ala Leu Ala Leu Arg Thr Ile Val Phe Ser Thr Ala Pro Ser Pro
20 25 30
Asn Arg Met Tyr Ser Tyr Lys Ser Ser Asp Arg Pro Arg Glu Trp Val
35 40 45
Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ile Pro Asp
15 50 55 60
Asp Glu Gln Lys Glu Phe Glu Asn Lys Ala Ala Ala Leu Ala Arg Ala
65 70 75 80
Ile Asp Leu Ala Glu Ser Ala Phe Leu Asp Lys Glu
85 90
20
<210> 103
<211> 616
<212> PRT
<213> *Nicotiana tabacum*

25
<400> 103
Met Gln Ser Leu Pro Ile Ser Tyr Arg Leu Phe Pro Ala Thr His Arg
1 5 10 15
Lys Val Leu Pro Phe Ala Val Ile Ser Ser Arg Ser Ser Thr Ser Ala
30 20 25 30
Leu Ala Leu Arg Val Arg Thr Leu Gln Cys Arg Cys Leu His Ser Ser
35 40 45
Ser Leu Val Met Asp Glu Asp Arg Phe Ile Glu Ala Ser Lys Ser Gly
50 55 60
35 Asn Leu Ile Pro Leu His Lys Thr Ile Phe Ser Asp His Leu Thr Pro
65 70 75 80
Val Leu Ala Tyr Arg Cys Leu Val Lys Glu Asp Asp Arg Glu Ala Pro
85 90 95
Ser Phe Leu Phe Glu Ser Val Glu Pro Gly Phe Arg Gly Ser Ser Val
40 100 105 110

109

Gly Arg Tyr Ser Val Val Gly Ala Gln Pro Ser Met Glu Ile Val Ala
 115 120 125
 Lys Glu His Asn Val Thr Ile Leu Asp His His Thr Gly Lys Leu Thr
 130 135 140
 5 Gln Lys Thr Val Gln Asp Pro Met Thr Ile Pro Arg Ser Ile Ser Glu
 145 150 155 160
 Gly Trp Lys Pro Arg Leu Ile Asp Glu Leu Pro Asp Thr Phe Cys Gly
 165 170 175
 Gly Trp Val Gly Tyr Phe Ser Tyr Asp Thr Val Arg Tyr Val Glu Asn
 10 180 185 190
 Arg Lys Leu Pro Phe Leu Arg Ala Pro Glu Asp Asp Arg Asn Leu Ala
 195 200 205
 Asp Ile Gln Leu Gly Leu Tyr Glu Asp Val Ile Val Phe Asp His Val
 210 215 220
 15 Glu Lys Lys Ala His Val Ile His Trp Val Gln Leu Asp Gln Tyr Ser
 225 230 235 240
 Ser Leu Pro Glu Ala Tyr Leu Asp Gly Lys Lys Arg Leu Glu Ile Leu
 245 250 255
 Val Ser Arg Val Gln Gly Ile Glu Ser Pro Arg Leu Ser Pro Gly Ser
 20 260 265 270
 Val Asp Phe Cys Thr His Ala Phe Gly Pro Ser Leu Thr Lys Gly Asn
 275 280 285
 Met Thr Ser Glu Glu Tyr Lys Asn Ala Val Leu Gln Ala Lys Glu His
 290 295 300
 25 Ile Ala Ala Gly Asp Ile Phe Gln Ile Val Leu Ser Gln Arg Phe Glu
 305 310 315 320
 Arg Arg Thr Phe Ala Asp Pro Phe Glu Val Tyr Arg Ala Leu Arg Ile
 325 330 335
 Val Asn Pro Ser Pro Tyr Met Thr Tyr Ile Gln Ala Arg Gly Cys Ile
 30 340 345 350
 Leu Val Ala Ser Ser Pro Glu Ile Leu Thr Arg Val Lys Lys Arg Arg
 355 360 365
 Ile Val Asn Arg Pro Leu Ala Gly Thr Ser Arg Arg Gly Lys Thr Pro
 370 375 380
 35 Asp Glu Asp Val Met Leu Glu Met Gln Met Leu Lys Asp Glu Lys Gln
 385 390 395 400
 Arg Ala Glu His Ile Met Leu Val Asp Leu Gly Arg Asn Asp Val Gly
 405 410 415
 Lys Val Ser Lys Pro Gly Ser Val Asn Val Glu Lys Leu Met Ser Val
 40 420 425 430

110

Glu Arg Tyr Ser His Val Met His Ile Ser Ser Thr Val Ser Gly Glu
435 440 445
Leu Leu Asp His Leu Thr Cys Trp Asp Ala Leu Arg Ala Ala Leu Pro
450 455 460
5 Val Gly Thr Val Ser Gly Ala Pro Lys Val Lys Ala Met Glu Leu Ile
465 470 475 480
Asp Gln Leu Glu Val Ala Arg Arg Gly Pro Tyr Ser Gly Gly Phe Gly
485 490 495
Gly Ile Ser Phe Ser Gly Asp Met Asp Ile Ala Leu Ala Leu Arg Thr
10 500 505 510
Met Val Phe Leu Asn Gly Ala Arg Tyr Asp Thr Met Tyr Ser Tyr Thr
515 520 525
Asp Ala Ser Lys Arg Gln Glu Trp Val Ala His Leu Gln Ser Gly Ala
530 535 540
15 Gly Ile Val Ala Asp Ser Asn Pro Asp Glu Glu Gln Ile Glu Cys Glu
545 550 555 560
Asn Lys Val Ala Gly Leu Cys Arg Ala Ile Asp Leu Ala Glu Ser Ala
565 570 575
Phe Val Lys Gly Arg His Lys Pro Ser Val Lys Ile Asn Gly Ser Val
20 580 585 590
Pro Asn Leu Phe Ser Arg Val Gln Arg Gln Thr Ser Val Met Ser Lys
595 600 605
Asp Arg Val His Glu Lys Arg Asn
610 615